AN IONOSPHERIC REMOTE SENSING METHOD USING AN ARRAY OF NARROWBAND VLF TRANSMITTERS AND RECEIVERS

A Dissertation
Presented to
The Academic Faculty

By

Nicholas C. Gross

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

December 2018

AN IONOSPHERIC REMOTE SENSING METHOD USING AN ARRAY OF NARROWBAND VLF TRANSMITTERS AND RECEIVERS

Approved by:

Dr. Morris Cohen, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Sven Simon
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. Mark Davenport
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Paul Steffes
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Mark Gołkowski
School of Electrical Engineering
University of Colorado Denver

Date Approved: September 7, 2018

To my parents

Theresa and Brian

and to my fiancé

Shraddha

ACKNOWLEDGEMENTS

I would first like to share my immense gratitude to my advisor, Professor Morris Cohen. He challenged me to pursue my own original ideas and was always available to give guidance and insight. His thoughtful mentorship and sound advice have been invaluable. I am also grateful to have had such an excellent undergraduate research advisor, Professor Mark Gołkowski. Thank you for introducing me to the VLF community and helping me build a strong foundation of electromagnetics and plasma physics knowledge.

Thank you to my other three thesis committee members: Professor Sven Simon for further developing my understanding of plasma physics, Professor Paul Steffes for the wonderful conversations about RF and planetary atmospheres, and Professor Mark Davenport for teaching me statistical signal processing techniques. I also owe a special thank you to Dr. Ryan Said for his time and guidance on MSK demodulation.

I am indebted to all the members of the Georgia Tech LF Radio Lab. This thesis would not exist if it were not for everyone's contributions toward the construction, deployment, operation, and maintenance of the Lab's radio receivers. Thank you to Jackson McCormick, Evan Worthington, Marc Higginson-Rollins, Nate Opalinski, Parker Singletary, Lee Thompson, and Nikhil Pailoor. You have all been tremendous friends.

I deeply appreciate the love and support my family has given me. I thank my sister, Katie, for her friendship and my parents, Theresa and Brian, for the countless sacrifices they have made. I finally thank my lovely fiancé, Shraddha, for her unwavering support and endless patience during my research endeavors.

This work was supported by the National Science Foundation under grants AGS 1451142 and AGS 1653114 (CAREER) to the Georgia Institute of Technology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Narrowband very low frequency (VLF) remote sensing has proven to be a useful tool for characterizing the ionosphere's D region (60 to 90 km altitude) electron density profile. VLF remote sensing experiments typically use a single narrowband VLF transmitter and receiver pair to determine a widely used two parameter exponential electron density profile, first introduced by *Wait and Spies* [1964] as waveguide parameters. Electron density profile inference with a single transmitter and receiver pair reveals temporal characteristics of the D region, however, more than one transmitter and receiver pair are needed to deduce spatial D region properties.

This work expands upon single transmitter and receiver electron density profile inference methods to create a more generalized narrowband VLF remote sensing method that concurrently resolves the two-parameter electron density profile along an arbitrary number of transmitter and receiver paths. A target function is constructed to take in a single time step of narrowband amplitude and phase observations from an arbitrary number of transmitter and receiver combinations and return the inferred waveguide parameters along all paths. The target function is approximated using an artificial neural network (ANN). Synthetic training data is generated using the US Navy's Long-Wavelength Propagation Capability (LWPC) program, which is then used to train the ANN. Real-world performance of the ANN is measured in two ways. First, ANN inferred waveguide parameters are compared to a variety of previously published narrowband VLF remote sensing experiments. Second, ANN inferred waveguide parameters are used in LWPC to predict narrowband VLF amplitude and carrier phase at a receiver that was withheld when performing the waveguide parameter inference. Results show the approximated target function performs well in capturing temporal and spatial characteristics of the D region.

CHAPTER 1
INTRODUCTION

1.1 Research Purpose

On the Earth's surface, the atmosphere is a neutral gas, but in the space environment near the Earth, where satellites and astronauts orbit, the environment instead consists mainly of charged electrons and ion species, which is known as a plasma. In between Earth's ground and near-Earth space exists a region where the neutral gas transitions to a plasma, and this transition region is called the ionosphere (60−500 km in altitude). For some technologies, like communications between ground stations and satellites, the ionosphere is a nuisance and degrades system performance. Other technologies, such as over-the-horizon (OTH) radar and submerged submarine communication, fundamentally rely on the ionosphere to serve as an electrically conductive boundary. Regardless of the ionosphere being an inconvenience or a necessity for system operation, systems that are affected by the ionosphere can improve their performance by better understanding the dynamics of this weakly to strongly ionized transition region. This thesis focuses on a specific type of radio wave communication, known as very low frequency (VLF 3−30 kHz) communication, and the interaction between VLF waves and the lowest portion of Earth's ionosphere, known as the D region (60−90 km altitude).

Electrical properties of the D region are challenging to measure. The altitude range of the D region is inaccessible for conventional *in situ* measurement techniques, such as aircrafts, balloons, and spacecraft. Sounding rocket measurements are a useful tool for directly measuring the ionosphere [*Seddon*, 1958; *Aikin et al.*, 1964; *Kane and Troim*, 1967], however these types of measurements are costly and cannot be maintained continuously nor over a large area. Ionospheric diagnostic methods, such as ionosondes, incoherent scatter radar, or monitoring total electron count from GPS signals, work well when inferring ionospheric properties above ~90 km, where the density of charged particles is much higher, but are often unreliable when examining the D region [*Rowe et al.*, 1974; *Sechrist*, 1974;

Mathews et al., 1982].

VLF radio waves have proven to be a useful diagnostic tool for remote sensing the D region. A VLF wave reflects nearly completely from the D region, unlike higher frequencies which pass through the D region. Upon reflection, the amplitude and/or phase of the VLF wave changes depending on the current physical state of the D region. Observations of these amplitude and phase changes can then be compared to a theoretical model to infer electrical properties of the D region; this technique is known as VLF remote sensing. The goal of this thesis is to develop a VLF remote sensing method, with an array of transmitters and receivers, to gain a deeper understanding of the complex dynamics of the D region and improve predictions for narrowband VLF propagation.

1.2 VLF Radiation Sources

VLF, and to a lesser extent Low Frequency (LF 30−300 kHz), radio waves efficiently reflect from the D region ionosphere and Earth's surface. Within the D region, the electron density, which dominates the electrical properties, is rapidly increasing with altitude. As a VLF wave propagates upward into the D region, it rapidly refracts due to this sharp gradient [*Budden*, 1985, p. 307]. In practice this refraction occurs over a narrow altitude range. Once an upward propagating VLF wave reflects from the D region, it begins to propagate downward towards Earth's surface, which is also a good conductor for VLF waves. These two concentric conducting sphere-like shells form what is commonly referred to as the Earth-ionosphere waveguide (EIWG), and allow VLF waves to efficiently propagate to global distances.

A simplified (not to scale) cartoon of VLF waveguide propagation is shown in Figure 1.1. The red arrow represents a single ray emanating from the transmitter and then reflecting between the surfaces of the D region ionosphere and Earth's surface. Note that the ray has a curve to it during its reflection with the ionosphere. This is done to emphasize that the wave does not abruptly reflect as if the D region is a perfect conductor, but instead the wave refracts and eventually heads back towards Earth. The reflection at Earth is dependent on the conductivity and permittivity at the point of reflection. Here, the ray reflects where

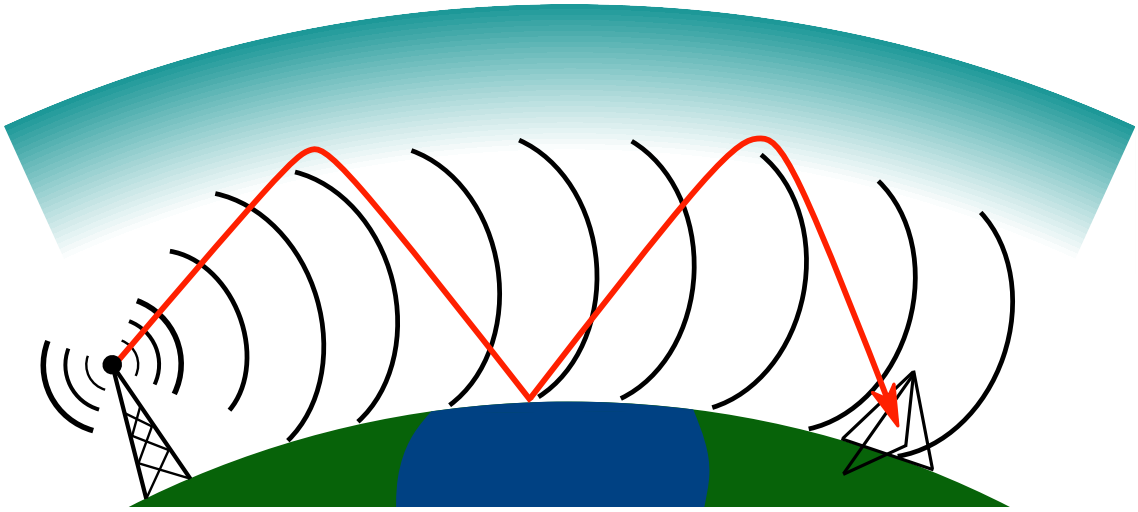

Figure 1.1: Simple EIWG propagation cartoon. The arrow shows a single ray reflecting off the D region and Earth's ground. The arcs represent a more realistic picture of the transmitter filling up the waveguide with VLF energy which then propagates towards a receiver.

there is water. If the water is salt water, which is a good conductor ($\sigma = 4$), then the wave would refract much quicker than it does in the ionosphere. A single ray isn't the complete picture, however, in practice a transmitter will illuminate the entire waveguide with VLF energy, and a portion of that energy will propagate within the EIWG, as depicted by black arcs propagating towards a receiver.

VLF radiation is ubiquitous throughout the world. The three most common VLF radiating sources are VLF transmitters, lightning, and electromagnetic noise generated from human-made devices and infrastructure, which will be referred to as electromagnetic interference (EMI) [*Chrissan and Fraser-Smith*, 1996]. Some examples of EMI are electromagnetic emissions from 50 and 60 Hz power lines (whose higher harmonics often reach into the VLF band), motors for heating ventilation and air conditioning (HVAC), and DC-to-AC converters often found in solar panels. For those who use sensitive instruments to measure VLF, it is almost always the case that a device that uses some form of alternating current will produce an appreciable amount of EMI in the VLF spectrum. With the exception of power line radiation, EMI typically does not efficiently couple into the Earth-ionosphere waveguide and is seen only when close to the EMI source (usually 10s of meters). VLF

transmitters and lightning, however, produce VLF energy that efficiently couples into the Earth-ionosphere waveguide and are the two most common sources for VLF EIWG propagation. Power line radiation can also be detected at long range [*Cohen et al.*, 2010b] and have also been detected in space [*Němec et al.*, 2015].

1.2.1 Narrowband VLF Transmitters

From 1971 to 1997, a global network of terrestrial VLF transmitters were operated by the US and six other nations, forming the *Omega* global navigation system [*McRae and Thomson*, 2000]. These transmitters were phase stable navigation beacons. VLF receiver operators would be able to discern their location by comparing the measured phase from multiple VLF transmitters. Since VLF waves propagate to global distances within the EIWG, only 8 VLF transmitters operating in the $10-14$ kHz frequency range were required to maintain global navigation coverage [*Swanson*, 1983]. Profound advancements in VLF transmitter design and VLF propagation modeling occurred in the 1950's and 1960's in preparation for the deployment of the *Omega* system, and VLF research from this time period laid the foundation for much of the VLF propagation theory used today. Advancements in global navigation satellite systems (GNSS) eventually led to the decommissioning of the *Omega* system.

Today, the US Navy and other countries operates VLF transmitters across the world to facilitate communication to submerged submarines. Some of these submarine transmitters are repurposed *Omega* transmitters. VLF transmitters are an obvious choice for submarine communication because global coverage can be obtained with a small number of transmitters, and VLF waves penetrate salt water much deeper than high frequency waves. Saltwater has a skin depth (a metric for measuring how well a wave penetrates a medium) of 1.6 m at 25 kHz and a skin depth of 0.0063 m at 1.575 GHz (L1 frequency for GPS), where skin depth is $\delta = 1/\sqrt{\omega\mu\sigma/2}$ for $\sigma/(\omega\epsilon) \gg 1$, relative permittivity (ϵ_r) is 81, and conductivity (σ) is 4 S/m [*Inan and Inan*, 2000, p. 56]. So, a 25 kHz wave has a skin depth that is three orders of magnitude larger than a 1.575 GHz GPS signal, which is an appreciable difference for a submarine that is trying to stay submerged and maintain covert operation.

There are two fundamental barriers to using VLF as a means for communication. First,

the bandwidth for VLF is limited; the entire VLF range is only 27 kHz wide, which is not much considering normal human speech uses ~3 kHz bandwidth. Nothing can be done about this innate bandwidth problem, and so the communication capacity of VLF is fundamentally limited to slow data transfer speeds. Second, VLF wavelengths are extremely long. A conventional quarter-wave monopole at a frequency of 10 kHz (~30 km wavelength) would need to stand 7.5 km in the air. Creating such a large antenna via conventional means is a challenging task, although some multi-kilometer antenna projects have proven successful [e.g., *Raghuram et al.*, 1974; *Carpenter et al.*, 1988; *Smith and Hanselman*, 1978]. With careful engineering, the antenna size problem can be mitigated, but at the cost of operational bandwidth [*Watt*, 1967].

Earth's ground is a good conductor in the VLF range, and so it follows from Maxwell's boundary conditions that a ground-based VLF antenna would excite VLF waves that have a vertically polarized electric field. A quarter-wave monopole would be a great antenna choice, except the physical height of the antenna is limited since towers usually extend to only 100's of meters tall. Instead, an antenna that is physically short but appears to be electrically tall is needed.

Consider a quarter-wave long transmission line that is terminated with a high impedance. Transmission line theory shows that the current through the load will be small but the voltage across the load will be large, which means the current and voltage are close to 90° out of phase with each other at the load, but at the source the current and voltage are exactly in phase with each other. A measurement of the current phase and voltage phase at some point towards the middle of the transmission line reveals that the phase of the current is leading the phase of the voltage. The idea is to physically reduce the length of the transmission line while still maintaining the electrical appearance of the original quarter-wave transmission line. If the length of the transmission line were reduced, then adding extra capacitance to the load will maintain the relationship of the current phase leading the voltage phase. Another way of saying this is that a shorter transmission line with the correct capacitance added to the load will have the same the input impedance as a quarter-wave transmission line.

Many VLF antennas use a method that is similar to this transmission line exercise,

known as *top-loading* [*Watt*, 1967]. A top-loaded VLF antenna will have a wire that runs vertically upward, using towers or mountains for support, and will add extra capacitance to the top of the antenna by stringing wires, which are connected to the cable at the top of the tower or mountain, horizontally above the ground. The collective series of horizontal wires at the top of the antenna is typically hundreds of acres in size and referred to as a *top-hat*. The top-hat and Earth's ground form the capacitance needed to make a physically short antenna appear electrically long. In practice, a top-loaded VLF antenna is complemented with a large series inductor (which usually has its own dedicated building) at the base of the feed wire. Together, the capacitance from the top-loading and the large inductor at the base create a high Q condition (quality factor, the ratio of the stored and dissipated energy) [*Swanson*, 1974]. This high Q condition limits the bandwidth of the antenna because the antenna becomes appreciably inductive with a small increase in frequency and appreciably capacitive with a small decrease in frequency, and so the operating frequency range is limited to a range of a few hundreds of Hz at best.

Power, voltage, and current are also problems for VLF antenna systems [*Gobbel*, 1967]. VLF antennas are inefficient and much of the input power goes into ground and copper losses, therefore VLF antennas operate with a high input power that sometimes exceeds 1 MW. Corona formation is usually the limiting factor for a VLF antenna's maximum power. High voltage regions on the antenna that are near ground potential and physically sharp points, such as bends or frays in a wire, are common places for corona formation. The radiation resistance of the antenna is small, on the order of 1 Ω, and thus a large amount of power being injected into the antenna results in large currents traveling along the cables. The antenna and electrical systems need to be engineered to handle these high power, voltage, and current conditions which makes the antennas more expensive and difficult to design.

1.2.2 Radio Atmospherics

Lightning strokes are excellent VLF radiators and complement narrowband VLF transmitters in different ways. Lightning is distributed all over Earth, whereas there are a small number of narrowband VLF transmitters permanently stationed throughout the world.

Lightning also generates broadband emissions that permeate the entire VLF band, whereas narrowband VLF transmitters have a very small bandwidth. The drawback for using lightning as a VLF source for ionospheric remote sensing is that the distribution of lightning is highly variable, and the VLF emissions are unique from one lightning flash to the next. About three quarters of lightning flashes are characterized as intracloud, meaning they occur entirely within a thundercloud. Most of the remaining flashes are called cloud-to-ground since they establish a conducting path to some location on the ground. Within cloud-to-ground lightning, there are positive and negative flashes, which refer to the region within the cloud from which the lightning discharge originated. Intracloud flashes are poor VLF radiators for two reasons. First, they tend to have lower peak currents [*Boccippio et al.*, 2001]. Second, the current travels mostly in a horizontal sense and much of the electric-field is canceled out by Earth's ground due to Maxwell's boundary conditions (an E-field horizontal to a perfect electric conductor will become zero at the boundary of the conductor) [*Smith et al.*, 2004].

A lightning flash is the entire cumulative discharge for a lightning event, and during a flash there are typically multiple unique discharges, known as strokes. It is each individual stroke that produces a sferic, and so a single flash can produce multiple sferics that are separated 10's of milliseconds apart.

A simple explanation for VLF radiation from a cloud-to-ground lightning stroke is that it is a quick movement of charge in a downward/upward direction over a distance that is on the same order of magnitude as VLF wavelengths. This vertical current channel, which can have peak currents of several hundred kA, is roughly analogous to a vertical monopole attached to a ground plane (Earth), although the horizontal currents within the thundercloud may also contribute to the VLF radiation. Hence, lightning radiates VLF energy efficiently and much of the energy couples into the Earth-ionosphere waveguide [*Taylor*, 1960].

Radio energy emitted from a lighting stroke is commonly referred to as a radio atmospheric, or *sferic*, and the exact radiation pattern and spectral energy of the sferic is usually not well known, due to the chaotic nature of lightning. Radiated sferic energy typically peaks at a frequency between 5 to 10 kHz and the power spectrum falls off inversely with frequency [*Le Vine*, 1987]. *Christian et al.* [2003] used the Optical Transient

Detector instrument aboard the MicroLab-1 satellite to estimate an average of ~45 lighting flashes/sec (includes intracloud and cloud-to-ground flashes) over the entire world. A high occurrence rate of lightning flashes produced around the world joined with the fact that sferics efficiently propagate in the Earth-ionosphere waveguide means that the natural VLF radio band is constantly filled with and dominated by energy originating from lightning [*Chrissan and Fraser-Smith*, 1996].

Source locations for sferic emissions can be accurately found using a lightning detection network. The National Lightning Detection Network (NLDN) is an excellent lightning detection network that operates in the US and can detect both cloud-to-ground and intracloud strokes with high stroke location accuracy [*Holle et al.*, 2016]. Detection of lightning on a global scale is done using the Vaisala Inc. GLD360 network [*Said et al.*, 2010, 2013], the World Wide Lightning Location Network (WWLLN) [*Rodger et al.*, 2009], or the Earth Networks Total Lightning Network (ENTLN) [*Liu and Heckman*, 2011]. These systems measure VLF/LF sferic emissions from the stroke and use time of arrival and magnetic direction-finding methods to determine the location of the stroke. Since the sensor coverage of these three systems is global and the density of network receivers is sparser than NLDN, the global networks have poorer location accuracy than NLDN.

1.3 Data Acquisition

The Georgia Tech LF Radio Lab has designed a VLF/LF/MF radio receiver, known as the LF AWESOME (Atmospheric Weather Electromagnetic System for Observation, Modeling, and Education) receiver [*Cohen et al.*, 2018a]. This receiver is an evolutionary design of the *Cohen et al.* [2010a] VLF AWESOME receiver. While the old version is named a VLF receiver and the new version is named an LF receiver, the LF AWESOME receiver functions just as well at VLF frequencies. Notable differences between the receivers are the LF version has a bandwidth of $0.5-470$ kHz, compared to the VLF receiver's $0.3-47$ kHz bandwidth, and the LF receiver has an (optional) integrated GPS module and analog-to-digital converter (ADC).

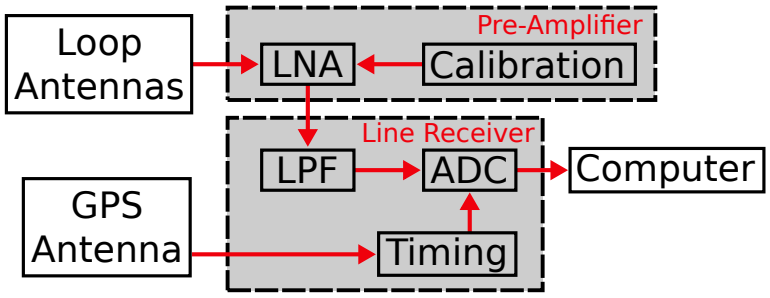

Figure 1.2: Block diagram of the LF AWESOME receiver.

1.3.1 Receiver Performance

Each LF AWESOME receiver is typically equipped with two magnetic field loop antennas (although a third can be used if desired), one aligned north and south (N/S) and the other east and west (E/W). A N/S aligned antenna is sensitive to transverse-magnetic (TM) waves propagating in the N/S direction, while an E/W antenna is sensitive to E/W traveling TM waves. Another common antenna type used in VLF receivers is a single vertical electric field monopole. The major advantage of using two loop antennas with the LF AWESOME receiver is the ability to perform direction finding (determining angle of arrival) and decompose the horizontal magnetic flux density into its azimuthal B_ϕ (magnetic flux density for TM waves) and radial B_r (magnetic flux density for transverse-electric (TE) waves is the superposition of B_r and B_z) components [*Krider et al.*, 1976].

Figure 1.2 shows a block diagram of the signal's path through the system. Directly next to the antenna is a low-noise amplifier (LNA) system that impedance matches to the loop antennas. A calibration circuit is embedded inside the pre-amplifier box and calibrates the system. Away from the antennas, inside a building or weather protected environment, are a line receiver and a computer. The pre-amplifier connects to the line receiver through a special audio cable that can be as long as 1,000s of feet. Having such a long cable allows for the loop antennas to be placed far away from buildings and sources of EMI, like 60 Hz power lines. At the line-receiver, the analog signal is converted to a digital signal and then sent to a computer.

Acquisition of the radio signal inside the line-receiver is accomplished using a 12^{th} order elliptical low pass filter (LPF) with a cutoff frequency of 470 kHz and a 16-bit analog-

to-digital converter (ADC) with a 1 MHz sampling rate. Data from the analog-to-digital converter is sent to the computer, where it is saved to disk and processed in real time. The MSK demodulation algorithm from *Gross et al.* [2018] (explained in Chapter 3) is used to extract the amplitude, carrier phase, and clock phase of each narrowband VLF transmitter on both the N/S and E/W channels. This real time information gathered from the demodulation algorithm is then sent to a server, named *Geniza*, operated by the LF Radio Lab on Georgia Tech's campus.

Precisely sampling the radio signal with a Universal Time (UT) aligned clock is of the utmost importance. Accurate comparison of VLF measurements between sites is dependent on each receiver, even those thousands of miles away, sampling the radio signal at the same moment. If each receiver samples the radio signal at a different time, then determining radio propagation characteristics through site comparison will not be possible. The LF AWESOME receiver uses the algorithm developed by *Cohen et al.* [2018a] to synchronize the sampling clock with GPS, and achieves a root-mean-square error (RMSE) of ~15 ns when compared to other reliable timing sources.

1.3.2 Transmitter and Receiver Array

Ten continuously operating LF AWESOME receivers are currently deployed within the US. Eight of the receiver locations are shown in Figure 1.3, and the final two reside in Alaska. Three VLF transmitters are located along the northern border of the continental US. These transmitters are constantly operating, except for a short time window every week where they go offline for maintenance, otherwise these three transmitters are used to communicate with submerged submarines. Long curves are shown that connect the transmitters and receivers, which are known as great circle paths (GCPs). A GCP is the shortest path between two points along the surface of a sphere. For a radio wave emanating from a transmitter, it is assumed that the radio wave will travel along its GCP. However, scattering objects, such as mountains or ionospheric perturbations, do cause multi-path VLF propagation and the assumption that a VLF wave strictly follows a GCP is not always a valid assumption.

The cumulative use of the transmitters and receivers in Figure 1.3 forms an array with an interesting geometry. Some GCPs intersect with other GCPs, while others have no

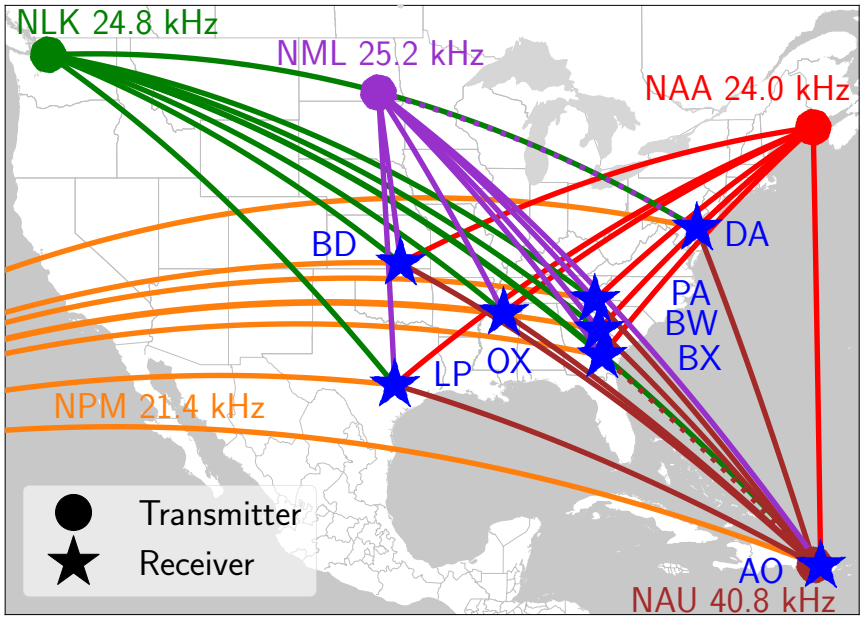

Figure 1.3: Map of narrowband VLF transmitters and Georgia Tech LF Radio Lab receivers.

intersections. The PA, BW, and BX receivers are located close to each other, and if the ground and ionosphere are considered to be slowly varying, then the electrical properties of Earth and the ionosphere along these close GCPs should be similar. Also, the frequency of the three northward transmitters are conveniently close to ~25 kHz. Therefore, the difference in transmitter frequency (and therefore, in the ionospheric response) is a minimal concern when comparing observations between these three transmitters.

1.4 VLF Waves in the Ionosphere

Earth's ionosphere is a plasma medium that efficiently reflects upward traveling VLF waves back towards Earth. A plasma is a mixture of three types of particles: neutrals, electrons, and ions. The approach taken here is to derive the complex valued index of refraction within the plasma medium, which largely defines the response of the ionosphere to a certain radio wave. This response allows us to derive how the wave will attenuate, shift in phase, or rotate polarization with each ionospheric reflection.

1.4.1 D Region Plasma Physics

During the daytime, radiation from the Sun dominates ionization in the D region, where major ionized constituents include: nitric oxide (NO) from Lyman-α (1216 Å) radiation, oxygen molecules (O_2) from X-ray and and Extreme UltraViolet (EUV) radiation, and nitrogen molecules (N_2) from EUV radiation [*Nicolet and Aikin*, 1960]. Lower levels of ionization are maintained during the nighttime by galactic cosmic ray radiation which ionizes NO, O_2, and N_2 [*Francey*, 1970].

Neutral particle densities are much larger than charged particle densities within the D region ionosphere. This means the D region ionosphere is a weakly ionized plasma. With the number of neutrals far exceeding the number of electrons and ions, it follows that most particle collisions will be between neutrals and either electrons or ions. The rate at which these collisions occur are known as electron-neutral collision frequency ν_{en} and ion-neutral collision frequency ν_{in}.

A VLF wave in Earth's lower atmosphere, near the ground, is for the most part unaffected by the neutral gas, and at this altitude a free space propagation assumption is a good approximation. We can therefore characterize the free space propagation of the VLF wave using the time-harmonic Ampere's law

$$\mathbf{k} \times \mathbf{B} = -\omega \mu_0 \epsilon_0 \mathbf{E} \tag{1.1}$$

and Faraday's law

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}, \tag{1.2}$$

where

j	is the imaginary unit ($j = \sqrt{-1}$),
ω	is the VLF wave radial frequency,
ϵ_0	is the permittivity of free space,
μ_0	is the permeability of free space,
$\mathbf{k}$	is the wave vector at frequency ω,
$\mathbf{E}$	is the electric field phasor vector at frequency ω,
$\mathbf{B}$	is the magnetic flux density phasor vector at frequency ω,

for low altitude propagation. This free space propagation assumption is not valid when the VLF wave is propagating within the ionosphere. Therefore, we seek to couple Maxwell's free space equations with plasma physics theory to understand VLF propagation outside and inside the ionosphere.

A common approach to solving plasma physics problems is to avoid the intractable method of accounting for each individual particle, and instead use macroscopic equations to simplify the analysis. This is done using a statistical mathematical statement, known as the Boltzmann equation, that accounts for forces, diffusion, and collisions for all particle species in the plasma. Macroscopic plasma equations are found by taking moments of the Boltzmann equation [*Inan and Golkowski*, 2011, Ch. 4]. The time-domain zeroth-order moment, known as the continuity equation, is

$$\frac{\partial N_\alpha(t)}{\partial t} + \nabla \cdot [N_\alpha(t)\ \mathbf{u}_\alpha(t)] = 0, \tag{1.3}$$

where

α	is an arbitrary charged particle species,
N_α	is the background particle density,
$\mathbf{u}_\alpha$	is the mean plasma fluid velocity vector.

The continuity equation is an expression that accounts for all particles within some differential volume. The change in particle population within the differential volume (left term) must be equal to the total number of particles that exited the volume (right term). This

equation assumes no particles are created or destroyed within the volume, but an additional term can be added to the continuity equation to account for this process if needed.

Taking the first-order moment of the Boltzmann equation gives the momentum transport equation

$$m_\alpha N_\alpha(t)\frac{\mathrm{d}\,\mathbf{u}_\alpha(t)}{\mathrm{d}\,t} = q_\alpha N_\alpha(t)\left[\mathbf{E}(t) + \mathbf{u}_\alpha(t) \times \mathbf{B}_0\right] - \nabla \cdot \boldsymbol{\psi}_\alpha(t) + \mathbf{S}_{\alpha\gamma}(t), \tag{1.4}$$

where

m_α	is the mass,
q_α	is the charge of the particle,
$\mathbf{B}_0$	is the Earth's static background magnetic flux density,
$\boldsymbol{\psi}_\alpha$	is the pressure tensor,
$\mathbf{S}_{\alpha\gamma}$	is the rate of change of momentum from collisions between two species.

Note that the amplitude of a radio wave in the plasma is assumed to be small; the implication of this assumption is that the electric field $\mathbf{E}(t)$ of the radio wave will be the dominant driver in perturbing the plasma medium and the magnetic flux density is negligible ($\mathbf{u}_\alpha(t) \times [\mathbf{B}(t) + \mathbf{B}_0] \approx \mathbf{u}_\alpha(t) \times \mathbf{B}_0$). Therefore, the only magnetic flux density in the momentum transport equation is Earth's static background magnetic flux density.

The momentum transport equation represents the balancing of forces within a differential volume. This equation states that Newton's second law of motion is balanced by three other force terms: Lorentz, pressure, and collision. The equation's left term comes from Newton's second law of motion (the product of mass and acceleration). The term $q_\alpha[\mathbf{E}(t) + \mathbf{u}_\alpha(t) \times \mathbf{B}_0]$ is the Lorentz force and it accounts for forces associated with the electric and magnetic fields. Pressure within the differential volume comes from random thermal particle motion and this pressure force is part of the term $\nabla \cdot \boldsymbol{\psi}_\alpha$. Species α may collide with another species γ and this collisional force is part of the term $\mathbf{S}_{\alpha\gamma}$.

Working with derivatives can be a tedious task. Often times it is useful to transform an equation from the time domain to the frequency domain to avoid time and spatial

derivatives. We have been using the (t) notation to represent a time varying variable, to represent these variables as phasors (a time-invariant complex value that represents a sinusoidal function) we will simply drop the (t) argument. The time-harmonic continuity equation is

$$j\omega N_\alpha - jN_\alpha \, \mathbf{k} \cdot \mathbf{u}_\alpha = 0, \tag{1.5}$$

where the spatial and temporal partial derivatives have been evaluated. The momentum transport equation has the one time derivative and becomes

$$m_\alpha N_\alpha j\omega \mathbf{u}_\alpha = q_\alpha N_\alpha (\mathbf{E} + \mathbf{u}_\alpha \times \mathbf{B}_0) - \nabla \cdot \boldsymbol{\psi}_\alpha - m_\alpha N_\alpha \nu_\alpha \mathbf{u}_\alpha \tag{1.6}$$

when written as a time-harmonic equation. Note the approximation $\mathbf{S}_{\alpha\gamma} \approx -m_\alpha N_\alpha \nu_\alpha \mathbf{u}_\alpha$ is used on the momentum transport equation. The term $\mathbf{S}_{\alpha\gamma}$ accounts for all collisions over the entire particle velocity distribution, but this distribution is difficult to empirically determine. Instead we approximate $\mathbf{S}_{\alpha\gamma}$ as a frictional damping term $(m_\alpha N_\alpha \nu_\alpha \mathbf{u}_\alpha)$, where ν_α is referred to as the effective collision frequency [*Inan and Golkowski*, 2011, p. 210].

Between equations (1.5) and (1.6), there is one term that is not accounted for, the pressure tensor. This problem arises from the moments of the Boltzmann equation, where taking the next moment introduce a new higher-order parameter. No matter how many moments of the Boltzmann equation are taken a new higher-order term is always introduced and a closed system of equations is never found. The approach here is to make some assumption that removes the highest-order term, the pressure tensor $\boldsymbol{\psi}_\alpha$. The pressure tensor is a function of the thermal velocity of a particle species, and an assumption is now made that the phase velocities of the VLF wave will be much larger than the *random* thermal motions of the particles, given that the amplitude of the wave is small [*Bittencourt*, 2004, p. 211]. Making this assumption means that the particles are relatively motionless compared to the phase velocity of the wave, thus the thermal velocity of the particles can be ignored. This has been found to be a good assumption for radio wave propagation in the D region [*Budden*, 1985, p. 4], but if it was not, higher-order moments of the Boltzmann equation could be included until a justifiable approximation is found. Using these higher-

order moments would result in the manifestation of plasma waves like Langmuir waves and ion-acoustic waves. By setting the $\boldsymbol{\psi}_\alpha$ term to zero we have effectively ignored particle temperature. This type of plasma is aptly named a *cold plasma* since the temperature is set to zero, and it is the foundation for magnetoionic theory (see *Ratcliffe* [1959] for an in-depth discussion).

It is useful to introduce two important parameters that help describe the plasma. Plasma frequency

$$\omega_{p\alpha} = \sqrt{\frac{N_\alpha q_\alpha^2}{m_\alpha \epsilon_0}} \tag{1.7}$$

is the natural resonant frequency of the particle species in the plasma, and gives an idea as to how quickly a plasma can react to a perturbation. As expected, the larger the mass of the particle the slower its response to a perturbation will be. Gyrofrequency, or cyclotron frequency,

$$\omega_{c\alpha} = \frac{-q_\alpha |\mathbf{B}_0|}{m_\alpha} \tag{1.8}$$

comes directly from the Lorentz force and is the frequency in which a species will gyrate around the background magnetic flux density $\mathbf{B}_0$. Depending on the species' charge q_α the gyrofrequency can be positive or negative and gives a sense of rotation (right-hand or left-hand) around $\mathbf{B}_0$. Using the fact that the current density vector $\mathbf{J}_\alpha$ of the species is the aggregate movement of charge in some direction

$$\mathbf{J}_\alpha = q_\alpha N_\alpha \mathbf{u}_\alpha, \tag{1.9}$$

and the cold plasma assumption ($\boldsymbol{\psi}_\alpha = 0$) on (1.6) gives an equation that relates the time-harmonic current density and electric field

$$j\omega \mathbf{J}_\alpha = \omega_{p\alpha}^2 \epsilon_0 \mathbf{E} - \omega_{c\alpha}(\mathbf{J}_\alpha \times \hat{\mathbf{b}}_0) + \nu_\alpha \mathbf{J}_\alpha \,, \tag{1.10}$$

where $\hat{\mathbf{b}}_0$ is the unit vector pointing in the $\mathbf{B}_0$ direction.

Rearranging the terms in (1.10) and assuming, without loss of generality, $\hat{\mathbf{b}}_0$ points in the $\hat{\mathbf{z}}$ direction produces an interesting relationship between the current density and the

electric field

$$\begin{aligned}\mathbf{J}_\alpha &= \boldsymbol{\sigma}_\alpha \mathbf{E} \\ &= \begin{bmatrix} \sigma_{\mathrm{P}\alpha} & -\sigma_{\mathrm{H}\alpha} & 0 \\ \sigma_{\mathrm{H}\alpha} & \sigma_{\mathrm{P}\alpha} & 0 \\ 0 & 0 & \sigma_{||\alpha} \end{bmatrix} \mathbf{E},\end{aligned} \tag{1.11}$$

where

$$\begin{aligned}\sigma_{||\alpha} &= \frac{\epsilon_0\, \omega_{p\alpha}^2}{(\nu_\alpha + j\omega)}, \\ \sigma_{\mathrm{P}\alpha} &= \frac{\epsilon_0\, \omega_{p\alpha}^2 (\nu_\alpha + j\omega)}{(\nu_\alpha + j\omega)^2 + \omega_{c\alpha}^2}, \\ \sigma_{\mathrm{H}\alpha} &= \frac{\epsilon_0\, \omega_{c\alpha}\, \omega_{p\alpha}^2}{(\nu_\alpha + j\omega)^2 + \omega_{c\alpha}^2},\end{aligned} \tag{1.12}$$

and the total current density is

$$\mathbf{J} = \sum_\alpha \boldsymbol{\sigma}_\alpha \mathbf{E} \tag{1.13}$$

[*Inan and Golkowski*, 2011, p. 164]. This relationship is also found in the simplified form of Ohm's law [*Chen*, 1984, p. 186]. From $\boldsymbol{\sigma}_\alpha$ it is apparent that the background magnetic field (recall that the background magnetic flux density $\mathbf{B}_0$ is assumed to be in the $\hat{\mathbf{z}}$ direction) causes anisotropy in the relationship between $\mathbf{J}_\alpha$ and $\mathbf{E}$. Electric fields perpendicular to $\mathbf{B_0}$ will produce currents both parallel and perpendicular to $\mathbf{E}_\alpha$ while still remaining perpendicular to $\mathbf{B}_0$, but electric fields parallel to $\mathbf{B}_0$ will produce currents also parallel to $\mathbf{B}_0$ that are proportional to $\mathbf{J}_\alpha$ by $\sigma_{||\alpha}$. As expected, setting $\mathbf{B}_0$ to zero ($\omega_{c\alpha} = 0$) forces $\sigma_{\mathrm{H}\alpha}$ to be zero and $\sigma_{\mathrm{P}\alpha}$ to become $\sigma_{||\alpha}$, which is the typical isotropic behavior of Ohm's law.

The conductivity tensor reveals multiple insights into the behavior of radio waves in a collisional, magnetized, cold plasma. First, the conductivity tensor is not isotropic (it is not a diagonal tensor and not all of its diagonal elements are equal), and thus the magnetized plasma exhibits anisotropic behavior. Second, all three conductivity parameters are complex, which means a radio wave will undergo attenuation as it propagates through the

medium if the complex value has a real component. This attenuation is due to charged particles colliding with neutral particles and is the primary mechanism for radio wave damping in the ionosphere. It is known that the ionosphere has a high neutral particle density [*Whitten and Poppoff*, 1971, p. 253], but if the neutral density goes to zero, causing the collision frequency ν_α to become zero, then the conductivity tensor's diagonal terms become purely imaginary, and attenuation along the diagonal of the matrix does not occur. Third, and perhaps less apparent, is the effect different species have on the total conductivity tensor ($\boldsymbol{\sigma} = \sum_\alpha \boldsymbol{\sigma}_\alpha$). Electrons have a mass much smaller than ions, and so a radio wave with a frequency much larger than the electron cyclotron frequency is dominated by electron effects (the ions moves so slow that it is as if they are motionless) [*Budden*, 1985, p. 55]. Under this circumstance, using only electrons for conductivity is a good approximation. However, radio waves lower in frequency, closer to the ion-plasma frequency and ion-gyrofrequency, are affected by both electrons and ions. Thus, multiple species must be accounted for in the tensor at lower frequencies (in the D region VLF waves below ~10 kHz may be noticeably affected by ions [*Cummer*, 1997]).

Ampere's law for free space is written in (1.1), but the conductivity tensor in (1.11) shows that the D region ionosphere cannot be treated as free space for VLF waves. Therefore, a unitless tensor

$$\boldsymbol{\epsilon}_p = \mathbf{I} - \frac{j\boldsymbol{\sigma}}{\epsilon_0\,\omega}, \tag{1.14}$$

where $\mathbf{I}$ is the identity matrix, is added to Ampere's law to account for the complicated permittivity of the plasma medium

$$\mathbf{k} \times \mathbf{B} = -\omega\mu_0\,\epsilon_0\,\boldsymbol{\epsilon}_p \cdot \mathbf{E}. \tag{1.15}$$

Combining (1.2) and (1.15) gives a wave equation

$$\mathbf{k}\,(\mathbf{k}\cdot\mathbf{E}) - (\mathbf{k}\cdot\mathbf{k})\,\mathbf{E} + \omega^2\mu_0\,\epsilon_0\,\boldsymbol{\epsilon}_p \cdot \mathbf{E} = \mathbf{0}, \tag{1.16}$$

which can also be written in a form that is reminiscent of an eigenvalue equation

$$\left[\left(\frac{c^2}{\omega^2}\mathbf{k}\mathbf{k}^{\mathrm{T}} + \boldsymbol{\epsilon}_p\right) - n^2\mathbf{I}\right] \cdot E = \mathbf{0}, \tag{1.17}$$

where $c = 1/\sqrt{\mu_0\,\epsilon_0}$ is the speed of light in free space and $n = kc/\omega$ is the index of refraction. If n^2 is treated as an eigenvalue, then the index of refraction is found by solving the characteristic equation

$$\det\left[\left(\frac{c^2}{\omega^2}\mathbf{k}\mathbf{k}^{\mathrm{T}} + \boldsymbol{\epsilon}_p\right) - n^2\mathbf{I}\right] = 0. \tag{1.18}$$

Solving the determinant, and assuming the radio wave frequency is sufficiently large so ion effects can be ignored, leads to the well-known Appleton-Hartree equation

$$n^2 = 1 - \frac{X}{1 - jZ - \frac{Y^2\sin^2\theta}{2(1-X-jZ)} \pm \sqrt{\left[\frac{Y^2\sin^2\theta}{2(1-X-jZ)}\right]^2 + Y^2\cos^2\theta}}\,, \tag{1.19}$$

where

$$X = \frac{\omega_{pe}^2}{\omega^2}\,, \qquad Y = \frac{\omega_{ce}}{\omega}\,, \qquad Z = \frac{\nu_e}{\omega}\,, \tag{1.20}$$

and θ is the spatial angle between the wavevector $\mathbf{k}$ and Earth's background magnetic flux density $\mathbf{B}_0$ [*Ratcliffe*, 1959, p. 19].

The Appleton-Hartree equation is rich with insights for radio wave propagation in a cold, collisional, and magnetized plasma. Many of these insights are not even applicable to D region VLF radio propagation and are more important for higher frequencies in higher regions of Earth's ionosphere. A few obvious relationships present themselves, though. First, the index of refraction, a key parameter for studying VLF radio wave propagation, is found if three things are known about the D region: magnetic flux density, electron-neutral collision frequency, and electron density. Notice that all these parameters vary as a function of altitude. Magnetic flux density is fairly simple to retrieve given the great magnetic models available for Earth (e.g., the World Magnetic Model [*Chulliat et al.*, 2014] and the International Geomagnetic Reference Field [*Thébault et al.*, 2015]). Finding the collision frequency and electron density is more difficult because existing models do not

predict the highly dynamic nature of these two parameters well. In fact, this thesis is mainly concerned with inferring the difficult to measure D region electron density profile through the use of VLF reflections. Second, the index of refraction is generally complex (the imaginary part of the refraction is associated with wave attenuation) and has some damping, but index of refraction becomes entirely real when the collision frequency goes to zero. This is exactly what is seen with the conductivity tensor in (1.11). Third, the angle of the background magnetic flux density, relative to the wave vector, plays a dominant role in controlling the index of refraction. This is why VLF wave propagation near Earth's equator (where $\mathbf{B}_0$ is close to parallel to Earth's ground) is very different than propagation near Earth's poles (where $\mathbf{B}_0$ is close to perpendicular to Earth's ground). Fourth, a $\pm$ sign exists in the equation, and it gives interesting polarization dependence for radio wave propagation. When the wave is propagating parallel to the background magnetic field ($\mathbf{k} \parallel \mathbf{B}_0,\ \theta = 0$) the '+' and '$-$' signs define left-hand and right-hand mode propagation, respectively. Waves propagating perpendicular to the background magnetic field ($\mathbf{k} \perp \mathbf{B}_0,\ \theta = \pm 90°$) propagate in either the ordinary mode ('+') or extraordinary mode ('$-$'). *Inan and Golkowski* [2011, Ch. 11] gives a good discussion on ordinary and extraordinary modes. One well known consequence of this polarization dependence is Faraday rotation for transionospheric radio waves at higher frequencies.

One final note about D region approximations. In a series of papers, *Wait and Walters* [1963a,b,c] showed the $\nu_\alpha \mathbf{J}$ term from (1.6) dominates over the $j\omega\mathbf{J}_\alpha$ term and can be set to zero when $\nu_e \gg \omega$ ($Z \gg 1$). Furthermore, the $\omega_{ce}(\mathbf{J}_e \times \hat{\mathbf{b}}_0)$ term can be neglected if $\nu_e \gg \omega_{ce}$ ($Z \gg Y$), making the plasma isotropic. If both of these assumptions are made then the complexity of the index of refraction in (1.19) greatly reduces to

$$n^2 = 1 - \sum_\alpha \frac{j\,\omega_{p\alpha}^2}{\omega\,\nu_\alpha}. \tag{1.21}$$

1.4.2 Wait and Spies Two-Parameter Model

Determining a high-fidelity D region electron density profile and electron-neutral collision frequency profile as a function of altitude is difficult. A simpler approach is to approximate

the two profiles using parameterized functions. *Wait and Spies* [1964] used laboratory measurements, ionospheric sounding rocket measurements, and VLF propagation measurements to empirically derive exponential profiles for the electron density and electron-neutral collision frequency. The exponential electron density profile

$$N_e(h) = 1.43 \cdot 10^{13} e^{-0.15h} e^{\beta(h-h')} \quad \mathrm{m}^{-3} \tag{1.22}$$

uses two parameters, h' km and β km^{-1}, to approximate the electron density of the D region. These two parameters are often referred to as waveguide parameters. An increase in h' may be thought of as the D region electron density moving upward (although not physically moving) in altitude (often associated with a reduction in D region ionization), while an increase in β implies that the gradient of the electron density profile has increased.

Collision frequency is an important property to understand because electron-neutral collisions are the principal effect in damping of VLF waves in the D region [*Bittencourt*, 2004, p. 410]. A simple exponential profile is used for the electron collision frequency

$$\nu_e(h) = \nu_0 e^{-\alpha h} \quad \text{collisions/sec}, \tag{1.23}$$

where ν_0 is the amplitude and α is the gradient for the collision frequency. A set of collisions frequency parameters, $\nu_0 = 1.82 \cdot 10^{11}$ collisions/sec and $\alpha = 0.15$ km^{-1}, were empirically determined by *Wait and Spies* [1964] and are typically used in modeling VLF propagation [*Cummer et al.*, 1998; *McRae and Thomson*, 2000; *Thomson et al.*, 2007; *Han and Cummer*, 2010a,b]. These *Wait and Spies* [1964] collision parameters are often not varied because changes in electron density affect VLF propagation much more dramatically than changes in collision frequency [*Budden*, 1985, p.12]. Instead, focus will be placed on inferring the waveguide parameters, h' and β.

Wait and Spies [1964] combined the electron density profile in (1.22) and the collision

	Summer	Winter
High Latitude	$\beta = 0.3$, $h' = 72$	$\beta = 0.3$, $h' = 72$
Middle Latitude	$\beta = 0.5$, $h' = 70$	$\beta = 0.3$, $h' = 72$

Table 1.1: Recommended profiles for h' and β during the daytime. Adapted from *CCIR* [1990].

Winter profiles					
	Magnetic dip range (degrees)	h'		β	
		f=10-30 kHz	f=30-60 kHz	f=10-30 kHz	f=30-60 kHz
High latitudes	90 to 75	76	76	$0.035f - 0.025$	1.2
Transition region	75 to 70	80	80	$0.035f - 0.025$	1.2
Middle latitude	< 70	87	88	$0.0077f + 0.31$	$0.0077f + 0.31$
Summer profiles					
All latitudes		$h' = 87$		$\beta = 0.0077f + 0.31$	

Table 1.2: Recommended profiles for h' and β during the nighttime. Adapted from *CCIR* [1990].

frequency profile in (1.23) to create an effective *conductivity parameter*

$$\begin{aligned} \omega_r &= \frac{\omega_{pe}^2(z)}{\nu_e(z)} \\ &= 2.5 \cdot 10^{13} e^{\beta(z-h')}. \end{aligned} \tag{1.24}$$

This conductivity parameter represents the height dependent terms in the simplified index of refraction in (1.21).

Morfitt [1977] and *Ferguson* [1980] performed extensive research into estimating general h' and β values for daytime and nighttime conditions at different latitudes. *CCIR* [1990] then simplified the estimated h' and β values from *Morfitt* [1977] and *Ferguson* [1980], and summarized them into two tables, Table 1.1 and Table 1.2. Note that these two tables are general estimates for h' and β and should be used only as a crude estimate or initial guess for the waveguide parameters. These tables do not take into account phenomenon such as the day-to-night and night-to-day terminator [*Crombie*, 1964], the solar cycle [*Thomson and Clilverd*, 2000], and transient D region perturbations [*Barr et al.*, 2000]. Interestingly, some frequency dependence for h' and β is seen in Table 1.2. This frequency dependence is discussed by *Morfitt* [1977] and *Ferguson* [1980] and is seen most often during winter nights. Although, a frequency dependence of the waveguide parameters does exist year-round and

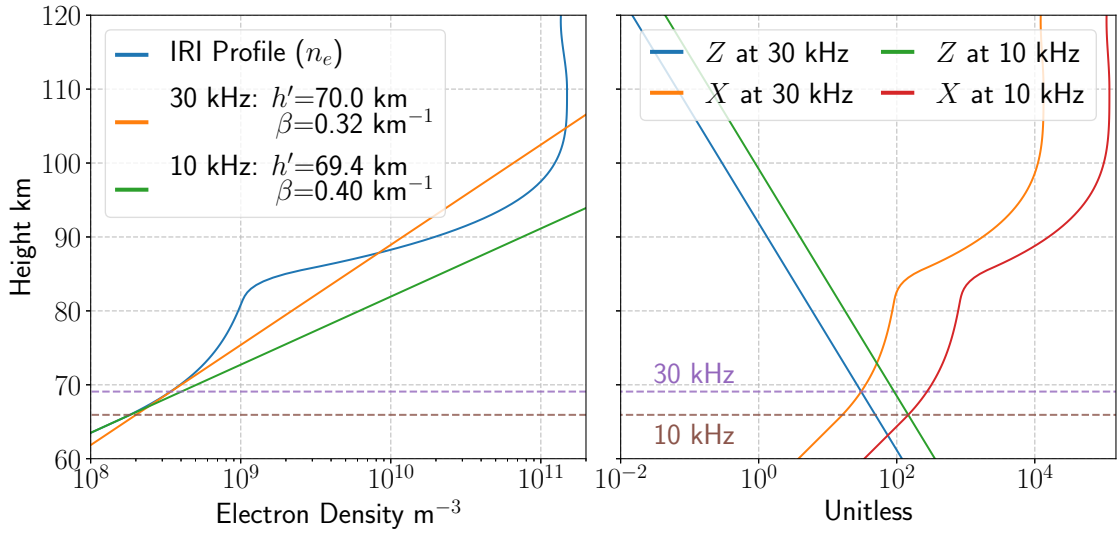

Figure 1.4: Example for estimating the effective reflection height for a 10 kHz and 30 kHz wave in an isotropic ionosphere. The h' and β values can be approximated by fitting a line to the reflection region of the IRI profile.

during the daytime but is usually less noticeable.

To conclude this section, an example VLF reflection that uses the Appleton-Hartree equation and waveguide parameters is given. Consider an isotropic ionosphere with no background magnetic field. If we take a daytime electron density profile from the International Reference Ionosphere (IRI) 2007 [*Bilitza and Reinisch*, 2008] and assume the *Wait and Spies* [1964] collision frequency, then we can estimate the effective reflection height for a VLF wave at a given frequency and infer the h' and β parameters. Figure 1.4 shows Z and X for a 10 kHz and 30 kHz VLF wave. Since a isotropic ionosphere is assumed, the intersection of the Z and X curves is a good approximation for the effective reflection height for each frequency [*Ratcliffe*, 1959, p. 92]. Once the effective reflection height is known, a line is fit to the IRI electron density profile reflecting region and the corresponding h' and β values are inferred. From the Z and X crossing, it is seen that the 10 kHz wave reflects at ~65.5 km and the 30 kHz wave reflects at ~69.0 km. The 10 kHz wave reflecting at a lower altitude than the 30 kHz wave follows a general rule: the lower in frequency a VLF or LF wave is, the lower in altitude it will reflect. An appreciable difference is seen between the inferred h' and β values for the two frequencies, where $\Delta h' = 0.6$ km and $\Delta\beta = -0.8$ km^{-1}.

1.5 Contributions

The following contributions to the field of VLF remote sensing are reported in this thesis:

- The utility of polarization ellipse analysis in narrowband VLF remote sensing applications is presented. It is shown that ellipse parameters, such as tilt angle and eccentricity, are important to consider when using VLF remote sensing techniques to infer properties of the D region ionosphere.

- A study involving radio wave scattering from conductive objects near VLF antennas is performed. It is found that conductive objects within 100s of meters of a VLF receiver's antenna can influence the phase observed at the antenna, and these phase deflections can exceed 10°. If not accounted for, this unexpected deflection in phase can have a non-negligible impact on VLF remote sensing techniques.

- A new method to concurrently solve for the average *Wait and Spies* [1964] two-parameter exponential profile ($\bar{h}'$ and $\bar{\beta}$) over multiple transmitter to receiver paths is developed. The method involves the approximation of a target function using an artificial neural network (ANN) and synthetically generated training data.

- A technique for evaluating the performance of the approximated target function (the ANN) is presented. Inferred $\bar{h}'$ and $\bar{\beta}$ values, from the ANN, are used to predict the amplitude and phase of a narrowband signal at a specific receiver. A ground truth error is then found by comparing this predicted amplitude and phase against the receiver's observed amplitude and phase. The advantage of this technique is the performance of the approximated target function is evaluated using ground truth measurements, whereas other $\bar{h}'$ and $\bar{\beta}$ inference methods use historical information to evaluate their performance, which can be less accurate because the ionosphere is highly dynamic and does not always follow historical trends.

1.5.1 Journal Publications

Gross, N. C., M. B. Cohen, R. K. Said, and M. Gołkowski (2018), Polarization measurements of VLF transmitters as an ionospheric diagnostic, *Journal of Geophysical Research - Space Physics*, 123 (1), 901-917, doi:10.1002/2017JA024907.

Cohen, M. B., **N. C. Gross**, M. A. Higginson-Rollins, R. A. Marshall, M. Golkowski, W. Liles, D. Rodriguez, J. Rockway (2018), Lower ionospheric response to the 2017 Great American Solar Eclipse, *Geophysical Research Letters*, 45 (8), 3348–3355.

McCormick, J. C., M. B. Cohen, **N. C. Gross**, and R. K. Said (2018), Spatial and temporal ionospheric monitoring using broadband sferic measurements, *Journal of Geophysical Research*, 123 (4), 3111–3130.

Cohen, M. B., R. K. Said, E. W. Paschal, J. C. McCormick, **N. C. Gross**, L. Thompson, M. Higginson-Rollins, U. S. Inan, J. Chang (2018), Broadband longwave radio remote sensing instrumentation, *Review of Scientific Instruments*, 89 (9).

Gołkowski, M., **N. C. Gross**, R. C. Moore, B. R. T. Cotts, and M. Mitchell (2014), Observation of local and conjugate ionospheric perturbations from individual oceanic lightning flashes, *Geophysical Research Letters*, 41 (2), 273-279, doi:10. 1002/2013GL058861.

1.5.2 Conference Presentations

Gross, N. C., and M. B. Cohen (2018), An Ionospheric Remote Sensing Method Using an Array of Narrowband VLF Transmitters and Receivers, Poster Presentation at 2018 AGU Fall Meeting; Washington, DC; paper SA23A-3165; December 10-14.

Gross, N. C., and M. B. Cohen (2017), Estimating Parameters for the Earth-Ionosphere Waveguide Using VLF Narrowband Transmitters, Poster Presentation at 2017 AGU Fall Meeting; San Francisco, CA; paper SA43A-2636; December 11-15.

Gross, N. C., and M. B. Cohen (2017), Estimation of the Earth-Ionosphere Waveguide Using Narrowband Transmitters, Oral Presentation at 2017 URSI GASS Radio Science Meeting; Montreal, Canada; abstract HGE16-3; August 21-25.

Gross, N. C., and M. B. Cohen (2017), Estimating the D-Region Ionospheric Electron Density Profile Using VLF Narrowband Transmitters, Oral Presentation at 2017 URSI US National Radio Science Meeting; Boulder, CO; abstract G2-3; January 4-6.

Gross, N. C., and M. B. Cohen (2016), Estimating the D-Region Ionospheric Electron Density Profile Using VLF Narrowband Transmitters, Poster Presentation at 2016 AGU Fall Meeting; San Francisco, CA; paper SA41A-2365; December 12-16.

Cohen M. B., **N. C. Gross**, F. G. Zoghzoghy, M. S. Briggs, M. Stanboro, and G. Fitzpatrick (2014), TGF Pulse and Radio Properties Detected at Close Range, Poster Presentation at 2014 AGU Fall Meeting; San Francisco, CA; paper AE31B-3394, December 15-19.

Gross, N. C., M. Gołkowski, R. C. Moore, B. R. T. Cotts, and M. Mitchell (2014), Local and Conjugate Ionospheric Disturbances from High Peak Current Oceanic Lightning Events, Oral Presentation at 2014 URSI US National Radio Science Meeting; Boulder; abstract EGH1-8; January 8-10.

Gross, N. C., M. Gołkowski, R. C. Moore, and B. R. T. Cotts, (2013), Local and Conjugate Ionospheric Disturbances from Rare High Peak Current Oceanic Lightning Events, Poster Presentation at 2013 AGU Fall Meeting; San Francisco, CA; paper AE33A-0321; December 9-13.

Gross, N. C., M. Golkowski, L. A. Barsikyan, and R. C. Moore (2013), Observed Local and Conjugate Ionospheric Disturbances from Rare High Peak Current Oceanic Lightning Events, Oral Presentation at 2013 APS Division of Plasma Physics Meeting; Denver, CO; abstract BAPS.2013.DPP.GO6.9; November 11-15.

CHAPTER 2
VLF PROPAGATION MODELING

The Earth-Ionosphere Waveguide (EIWG), the structure that allows VLF waves to propagate over the horizon to global distances, is a complicated and dynamic structure. It is bounded by the Earth's ground on the bottom side, and the ionosphere on the top side. Earth's ground has spatially diverse electrical properties but is for the most part static. Chemical properties of the ground, such as the salinity of the water or the elements in the soil, define these electrical properties and are accounted for with some degree of accuracy [*Wait and Spies*, 1965]. On the other hand, the Earth's ionosphere varies by location and is constantly changing, hence the electrical properties of this upper atmospheric region are difficult to account for. Some of the changes are repeatable and systematically predictable, like diurnal, seasonal, and solar cycle variations. Disturbances to the ionosphere, from a variety of transient geophysical events, can add shifts in the ionospheric characteristics that may be focused in one small area or spread over half the globe. Superimposed on top of that is a set of quasi-random variation in space in time that are difficult if not impossible to characterize.

Many different approaches have been used to model VLF waveguide propagation. *Lehtinen and Inan* [2008] present a model using the full-wave finite element method; *Chevalier and Inan* [2006] and *Marshall and Inan* [2010] use finite difference frequency domain (FDFD) methods; and *Thiel and Mittra* [1997], *Thevenot et al.* [1999], and *Hu and Cummer* [2006] each developed finite difference time fomain (FDTD) based models. A popular EIWG modeling method is to find the dominant resonant modes in the waveguide and then relate them to the total propagated signal. One of the most common VLF mode-finding programs is called Long-Wavelength Propagation Capability (LWPC) [*Ferguson*, 1998], and it will be the focus of this chapter and utilized heavily in subsequent chapters.

Much of the theory used in LWPC was developed between the 1950's and 1980's. Throughout the years, LWPC has undergone many iterations and changes in name. With

so many different versions and changes made to LWPC, it is difficult to pin down what theory is being used in the program and what is not. The emphasis of this chapter is to connect the literature that spans over these four decades and give high-level insight towards the inner workings of the LWPC program.

This chapter begins with a basic discussion of parallel-plate waveguide propagation theory, which is an idealized version of the EIWG. Afterwards, the main components to the LWPC program are discussed: reflection coefficients (1950's and 1960's), mode finding (1950's to 1970's), and mode conversion (1960's to 1980's). The chapter is concluded with some example LWPC runs and a discussion of past works and their use of propagation models to infer ionospheric characteristics from VLF observations.

2.1 Ideal Waveguide Theory

Consider the ideal parallel-plate waveguide shown in Figure 2.1. The bold arrows represent a single ray (electromagnetic plane-wave) reflecting back and forth within the waveguide. Dashed lines are drawn to represent the phase fronts of the wave (i.e., everywhere the phase of the wave is 0° (red) or 180° (blue) at a specific point in time). The wave is propagating in mode m with a wave vector $\mathbf{k}$ and is bounded by two perfect electric conducting plates, with infinite area in the $\hat{\mathbf{y}}$ direction, separated by distance a. When the wave reaches a conducting boundary, it reflects by angle $\theta_m \in \mathbb{C}$ and the $\hat{\mathbf{z}}$ component of $\mathbf{k}$ changes polarity. From the figure, it is seen that the wave varies by $\frac{\partial}{\partial x} = -jk\sin\theta_m$ in the $\hat{\mathbf{x}}$ direction and has no variation in the $\hat{\mathbf{y}}$ direction ($\frac{\partial}{\partial y} = 0$). This is the coordinate system that will be used for the rest of this thesis, it will always be assumed that the radiating source exists at $x = 0$ and the propagating wave will travel along the waveguide in the $\hat{\mathbf{x}}$ direction.

Discrete mode (resonant) solutions exists for the propagation of a wave in this ideal parallel-plate waveguide. The waveguide can be thought of as a resonator in the $\hat{\mathbf{z}}$ direction, and resonance will occur when the propagating wave produces a standing wave. A standing wave exists in $\hat{\mathbf{z}}$ direction when the distance between reflections in the $\hat{\mathbf{k}}$ direction is exactly

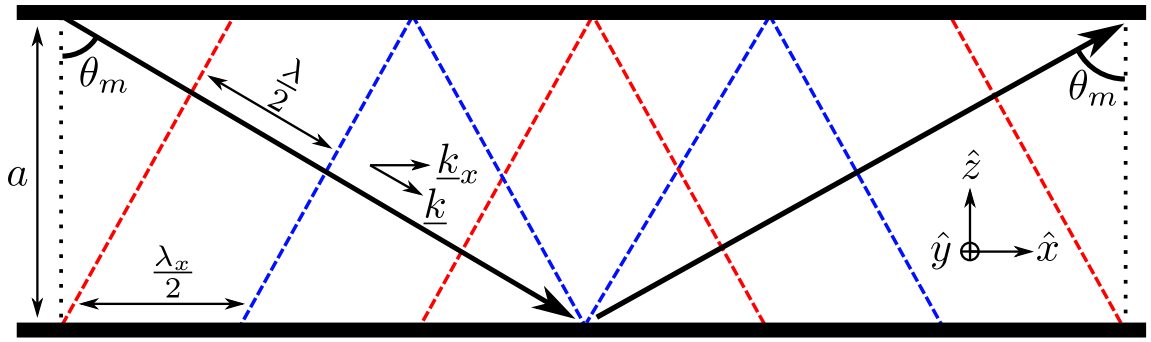

Figure 2.1: Ideal parallel-plate waveguide showing a single ray reflecting off of perfect conducting boundaries. The red and blue dashed lines represent phase fronts that are 180° out of phase from each other.

an integer multiple of $\lambda/2$, where λ is the free space wavelength, giving the relationship

$$a \cos\theta_m = \frac{m\lambda}{2} \quad \text{for} \quad m = 1, 2, 3, \ldots \tag{2.1}$$

[*Johnk*, 1988, p. 421]. Another way to think of this is as follows: the phase of the wave at any location must be equal to phase of the wave after it has reflected once from each boundary and then returned to the original altitude. Note that the θ_m solutions for (2.1) are all real valued, but θ_m is generally complex valued when the waveguide has lossy reflection boundaries, as is the case for the EIWG. Equation (2.1) is bounded to solutions that satisfy $\frac{m\lambda}{2a} \leq 1$, implying the waveguide is able to support less modes as the wavelength becomes longer. Waveguide cutoff occurs when $m\lambda/2 = a$, and waves with frequencies below the waveguide cutoff are unable to propagate within the waveguide. As the frequency approaches waveguide cutoff, the angle of reflection also decreases, which in turn means the wave will reflect more often and take a longer amount of time to propagate some distance down the waveguide in direction $\hat{\mathbf{x}}$. This excess propagation time, or delay, is equivalent to a decrease in group velocity

$$v_g = \frac{\omega\lambda}{2\pi} \sin\theta_m \tag{2.2}$$

and an increase in phase velocity

$$v_p = \frac{\omega\lambda}{2\pi \sin\theta_m} \tag{2.3}$$

for the $\hat{\mathbf{x}}$ direction. Therefore, dispersion becomes readily apparent within the waveguide as the frequency approaches cutoff.

Examination of a single propagating ray simplifies the analysis for parallel-plate waveguide theory, but in practice instruments measure the superposition of a large number of rays and multiple modes. Consider a hypothetical isotropically radiating (or omnidirectional) antenna placed inside the waveguide which transmits at a frequency well above cutoff. This transmitter will illuminate the waveguide with many modes in all directions. Instead of describing the infinite number of modes individually, a compact expression for the propagating modes may be derived, such as the magnetic flux density $\mathbf{B}$ in the $\hat{\mathbf{y}}$ direction

$$\mathbf{B} \cdot \hat{\mathbf{y}} = C \cos\left(\frac{m\pi}{a} z\right) e^{-j(\mathbf{k}_x \cdot \hat{\mathbf{x}})x}, \tag{2.4}$$

where C is a constant [*Johnk*, 1988, p. 422]. Note that this equation and all subsequent equations in this in chapter are written in time-harmonic (phasor) form. Modes within a parallel-plate waveguide are decomposed into two types: transverse electric (TE) modes and transverse magnetic (TM) modes. There is also a transverse electromagnetic (TEM) mode, though this can also be thought of as a special case of TM where the mode number m is 0. An electromagnetic wave with an electric field that is parallel to the $\hat{\mathbf{y}}$ axis, in Figure 2.1, is called a TE wave. Conversely, a wave with the magnetic field that is parallel to the $\hat{\mathbf{y}}$ axis, such as (2.4), is considered to be a TM wave. A TEM wave is equivalent to a uniform plane wave that does not reflect off either boundary in the parallel-plate waveguide.

An ideal parallel-plate waveguide is a good starting point for understanding VLF wave propagation. The previously discussed topics, such as waveguide cutoff, dispersion, and angle of reflection are relevant when discussing EIWG propagation. However, the theory quickly becomes more complicated when non-ideal behaviors are considered: imperfect reflecting boundaries, anisotropic behavior caused by Earth's background magnetic field, curvature of the Earth, multipath, variations in distance between the reflecting boundaries, etc. The next sections will expand upon the ideal waveguide theory and introduce some of these non-ideal behaviors.

2.2 Waveguide Reflections

In the previous section, the angle of reflection from (2.1) is found using simple geometry and the fact that the waveguide resonates in the $\hat{\mathbf{z}}$ direction, but this assumed that the boundaries were perfect reflecting conductors. Now consider the case were the boundaries are general conductive media and have complex valued upper $R_u(\theta)$ and lower $R_\ell(\theta)$ reflection coefficients, that change according to the angle of incidence $\theta \in \mathbb{C}$. From the Fresnel equations, we generally expect some of the wave's power to penetrate the boundary and the rest of the power to internally reflect in the waveguide. The amount of power that penetrates and reflects is dependent on the media and the angle of incidence. The fact that the waveguide resonates in the $\hat{\mathbf{z}}$ direction still holds, and so the relationship

$$R_u(\theta_m)e^{-jak\cos\theta_m}\, R_\ell(\theta_m)\, e^{-jak\cos\theta_m} = 1, \tag{2.5}$$

where $k = 2\pi/\lambda$ is the free space wave number, must be true [*Galejs*, 1972, p. 10]. This equation is a mathematical statement about the definition of a mode. The term $R_u(\theta_m)$ includes what happens to the wave phasor upon reflection from the ionosphere, similarly the term $R_\ell(\theta_m)$ describes the change to the phasor with respect to the ground reflection. The $a\cos\theta_m$ term in the two identical exponential terms $e^{-jak\cos\theta_m}$ account for free space propagation in the upward and downward directions.

Rearranging terms in (2.5) and using the identity $1 = e^{j2\pi m}$ gives

$$R_u(\theta_m)R_\ell(\theta_m)e^{-j2ak\cos\theta_m} = e^{j2\pi m}, \tag{2.6}$$

which is similar to the $ak\cos\theta_m = \pi m$ from (2.1). Note that the reflecting boundaries may be lossy in which case the real part of the reflection coefficient will be less than unity $\Re\{R(\theta)\} \leq 1$, but the magnitude of the complex reflection coefficient by definition remains unity. Thus there is some reduction in wave amplitude from the reflection, but this is compensated by the imaginary component of the reflection coefficient and $|R(\theta)| = 1$ remains true [*Budden*, 1961, p. 125].

2.2.1 Ionospheric Reflection Coefficient Matrix

In simpler waveguides, the reflection coefficient is a scalar (complex) value. However, Earth's magnetic field adds anisotropy to ionospheric reflections of VLF radio waves, which in turn causes coupling between TE and TM modes. Since the two modes are coupled together and can no longer be treated as two separable equations, it is typically said that these waves propagate in quasi-transverse magnetic (QTM) and quasi-transverse electric (QTE) modes. A complex valued reflection matrix, instead of a single scalar value, must be used to account for the coupling between these two distinct polarization types

$$\mathbf{R}(\theta) = \begin{bmatrix} {}_{\parallel}R_{\parallel}(\theta) & {}_{\perp}R_{\parallel}(\theta) \\ {}_{\parallel}R_{\perp}(\theta) & {}_{\perp}R_{\perp}(\theta) \end{bmatrix}, \tag{2.7}$$

where the $\parallel$ and $\perp$ specify whether it is parallel or perpendicular to the plane of incidence (the plane perpendicular to $\hat{\mathbf{z}}$), the prescript denotes the polarization for the incident wave, and the subscript is for the polarization of the reflected wave. The relationship between the incident $\mathbf{E}^i$ and reflected $\mathbf{E}^r$ electric field is then

$$\begin{bmatrix} E^R_{\parallel} \\ E^R_{\perp} \end{bmatrix} = \begin{bmatrix} {}_{\parallel}R_{\parallel}(\theta) & {}_{\perp}R_{\parallel}(\theta) \\ {}_{\parallel}R_{\perp}(\theta) & {}_{\perp}R_{\perp}(\theta) \end{bmatrix} \begin{bmatrix} E^I_{\parallel} \\ E^I_{\perp} \end{bmatrix}. \tag{2.8}$$

A relationship between the incident and reflected electric field is not enough to determine the reflection coefficient. Two polarized waves, orthogonal to each other, are required to solve for all four reflection matrix coefficients. Using the same coordinate system from Figure 2.1, it is found that

$${}_{\parallel}R_{\parallel} = \frac{H^R_y}{H^I_y}, \qquad {}_{\parallel}R_{\perp} = \frac{E^R_y}{\eta_0 H^I_y} \tag{2.9}$$

when $E^I_y = H^I_x = 0$, and

$${}_{\perp}R_{\parallel} = \frac{\eta_0 H^R_y}{E^I_y}, \qquad {}_{\perp}R_{\perp} = \frac{E^R_y}{E^I_y} \tag{2.10}$$

when $E^I_x = H^I_y = 0$, where H is the magnetic field and η_0 is the intrinsic impedance of free

space [*Budden*, 1985, p. 298]. It is also important to recognize that incident and reflected fields change the closer or further away they are from the boundary. Therefore, a reflection coefficient will take on different values depending on the distance from the boundary the electric and magnetic fields are observed at, and thus the reflection coefficients must be referenced to some distance from the boundary.

2.2.2 Ground Reflection Matrix

Earth's ground is the simpler of the two reflection matrices to solve for. We assume that the ground is isotropic and homogeneous over a region much larger than a wavelength. Therefore, no coupling between the TE and TM modes occurs during ground reflections and the downward looking ($-\hat{\mathbf{z}}$) reflection matrix simplifies to a diagonal matrix

$$\overline{\mathbf{R}}_g(\theta) = \begin{bmatrix} {}_{\parallel}\overline{R}^g_{\parallel}(\theta) & 0 \\ 0 & {}_{\perp}\overline{R}^g_{\perp}(\theta) \end{bmatrix}, \tag{2.11}$$

where the g means the altitude of the reflection matrix is taken to be at the ground. It has been shown that anisotropy does exist in the ground and is non-negligible for some regions [*Galejs*, 1972, p. 222], but it is often ignored anyway. It can be included via this formulation by simply expanding $\overline{\mathbf{R}}_g(\theta)$ to be non-diagonal.

A homogeneous ground means there is no $\hat{\mathbf{z}}$ dependence in the ground, and the reflection is simply a two media boundary problem, with the top media being free space. The classic Fresnel equations are used to solve this boundary problem

$$\begin{aligned} {}_{\parallel}\overline{R}^g_{\parallel}(\theta) &= \frac{n_g^2\cos\theta - \sqrt{n_g^2 - \sin^2\theta}}{n_g^2\cos\theta + \sqrt{n_g^2 - \sin^2\theta}} \\ {}_{\perp}\overline{R}^g_{\perp}(\theta) &= \frac{\cos\theta - \sqrt{n_g^2 - \sin^2\theta}}{\cos\theta + \sqrt{n_g^2 - \sin^2\theta}}, \end{aligned} \tag{2.12}$$

where the square ground index of refraction is

$$n_g^2 = \epsilon_r - \frac{j\sigma}{\omega}, \tag{2.13}$$

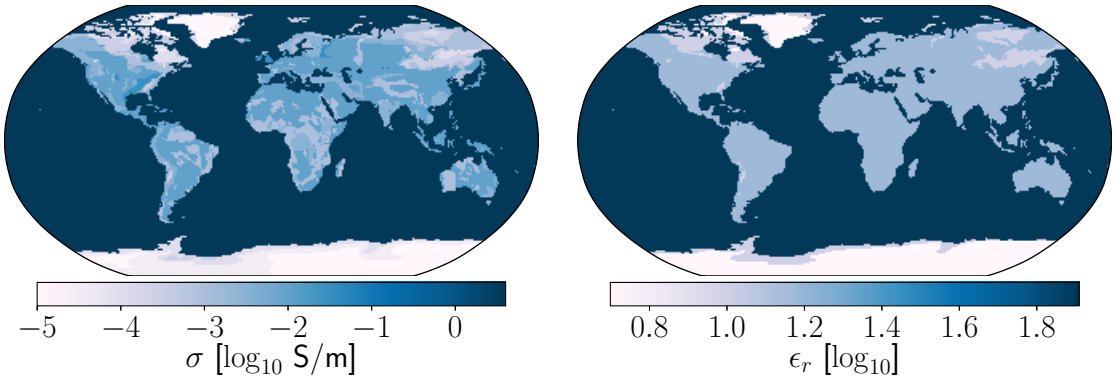

Figure 2.2: Map of Earth's conductivity (left) and relative permittivity (right).

ϵ_r is the relative permittivity of the ground, and σ is the conductivity of the ground.

Regions with old crystalline rocks, significant permafrost, and ice caps have some of the lowest conductivities and the most profound effects on VLF waveguide propagation. Greater depths of the ground's profile must be considered for lower conductivities, and these depths can exceed more than 50 meters for conductivities less than 10^{-2} S/m [*Morgan*, 1968]. Conductivity and relative permittivity profiles are extracted from the LWPC program and shown in Figure 2.2. It is seen that regions with deep permafrost and icecaps have the lowest conductivity and relative permittivity, and the oceans have the highest and are the best reflectors. Other interesting features are seen in the maps, such as the high conductivity along the Appalachian Mountains with an abrupt decrease in conductivity around the state of Georgia. These maps are used to determine reflection matrices (2.11) along a great-circle path from the transmitter to receiver.

2.2.3 The Booker Quartic

To find the ground reflection matrix, we simplified the problem by assuming an isotropic and homogeneous ground, but these assumptions do not hold for general ionospheric reflections. Recalling our discussion of the Appleton-Hartree equation (1.19), we assume the ionosphere is anisotropic and horizontally stratified (varies in the $\hat{\mathbf{z}}$ direction). Our goal is to determine the relationship between an obliquely upward traveling wave and its reflected downward traveling counterpart.

One way to simplify the reflection problem is to assume the ionosphere is a sharp reflecting boundary, with free space on one side and a dense homogeneous ionosphere on the other. This simplification has been used in multiple works to simplify the mathematics and computational burden [e.g., *Crombie*, 1961; *Johler and Walters*, 1960; *de Lisle*, 1967].

Sheddy [1968] formulated a closed general reflection matrix solution for a sharp transition between free space and a homogeneous ionosphere, and that solution is considered here because it has special application to the LWPC program. An upward traveling plane wave in free space that is incident upon a sharp ionospheric boundary will generally refract into two plasma modes, the ordinary mode and extraordinary mode. The method from *Sheddy* [1968] starts by solving for these two upward traveling plasma waves in the ionosphere and then uses boundary conditions to determine the initial upward traveling wave in free space. With the waves solved for in both medias, the reflection matrix is then found using a method similar to that in (2.9) and (2.10).

The susceptibility matrix $\mathbf{M}$ is useful for relating the electric displacement $\mathbf{D}$ and the electric field $\mathbf{E}$ through

$$\mathbf{D} = \epsilon_0(\mathbf{I} + \mathbf{M})\mathbf{E}, \tag{2.14}$$

where $\mathbf{I}+\mathbf{M}$ is the same as the permittivity matrix from (1.14), except here the background magnetic field is not assumed to be in a particular direction. Using a similar approach as before, a susceptibility matrix that accounts for the cold, collisional, anisotropic plasma properties of the ionosphere is found to be

$$\mathbf{M} = \frac{-X}{U(U^2 - Y^2)} \begin{bmatrix} U^2 - Y_x^2 & -jUY_z - Y_xY_y & jUY_y - Y_xY_z \\ jUY_z - Y_xY_y & U^2 - Y_y^2 & -jUY_x - Y_yY_z \\ -jUY_y - Y_xY_z & jUY_x - Y_yY_z & U^2 - Y_z^2 \end{bmatrix}, \tag{2.15}$$

where $U = 1 - jZ$ and $[Y_x, Y_y, Y_z] = -\mu_0 q_\alpha / m_\alpha [\mathbf{B}_0 \cdot \hat{\mathbf{x}},\ \mathbf{B}_0 \cdot \hat{\mathbf{y}},\ \mathbf{B}_0 \cdot \hat{\mathbf{z}}]$. Multiple charged species in the ionosphere are accounted for in this susceptibility matrix by taking the sum of the species dependent susceptibility matrices, $\mathbf{M} = \sum_\alpha \mathbf{M}_\alpha$. From Ampere's Law (1.1)

and Faraday's Law (1.2), it is known that

$$\begin{aligned} \boldsymbol{\Gamma}(\eta_0 \mathbf{H}) &= -(1+\mathbf{M})\mathbf{E} \\ \boldsymbol{\Gamma}\mathbf{E} &= \eta_0 \mathbf{H}, \end{aligned} \tag{2.16}$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0 & -q & 0 \\ q & 0 & -\sin\theta \\ 0 & \sin\theta & 0 \end{bmatrix}, \tag{2.17}$$

q is a solution to a quartic equation for wave propagation in the ionospheric region (note that q here is different than the electric charge of a species), and the free space plane wave normal is assumed to be in the xz-plane [*Budden*, 1985, p. 145]. Combining the two equations in (2.16) and setting $\mathbf{H} = \mathbf{J} = 0$, results in a relationship between the dispersion matrix $\mathbf{A}$ and the electric field

$$\begin{aligned} \mathbf{0} &= \left(\boldsymbol{\Gamma}^2 + \mathbf{I} + \mathbf{M}\right)\mathbf{E} \\ &= \mathbf{A}\ \mathbf{E} \end{aligned} \tag{2.18}$$

[*Sheddy*, 1968]. For $\mathbf{E}$ to have more than just the trivial solution, the determinant of the dispersion matrix must zero. Solving for the determinant

$$\det[\mathbf{A}] = 0, \tag{2.19}$$

leads to a fourth-order equation known as the Booker quartic,

$$q^4 + a_3 q^3 + a_2 q^2 + a_1 q + a_0 = 0, \tag{2.20}$$

where q is the unknown, and the coefficients a_i are defined by $\mathbf{M}$ and θ [*Booker*, 1936]. This fourth-order equation has four wave solutions $\{q_1, q_2, q_3, q_4\}$, describing ordinary and extraordinary modes traveling upwards and downwards. Upward traveling waves from the Booker quartic solution are used with the known boundary conditions to determine the

original free space incident wave; the other two downward traveling waves are not needed and ignored. Each field component of the four plasma waves will have a factor $\exp[-jk(x\sin\theta + qz)]$. *Sheddy* [1968] gives the final details for finding the reflection matrix $\mathbf{R}_h(\theta)$, at the boundary altitude h.

A reflection matrix from a sharply bounded ionosphere may be a good approximation for some cases, but we are also concerned with electron density distribution as a function of altitude. Thus, we want to find a method that determines the ionospheric reflection of a VLF wave that also accounts for horizontal stratification of the ionosphere.

Consider an ionosphere that is layered with discrete homogeneous slabs of thickness δz, and each successive layer is labeled with the next consecutive integer $i = 1, 2, 3, \ldots$ For each slab, solutions to the Booker quartic are found. This is equivalent to finding the discrete wave solutions every δz step in altitude. It is also known that fields of each of the four waves (solutions to the Booker quartic) will have a field component in the form of

$$F_i(z) = F_i e^{-jk(x\sin\theta + q[i]z)}, \tag{2.21}$$

where F_i is field and wave dependent at slab i [*Budden*, 1961, p. 103]. We wish to generalize $F_i(z)$ into a continuous function that bridges across all slabs $F(z)$, and we next consider techniques to do that.

2.2.4 The WKB Approximation

A common approach to handling the height variation in electron density is to apply the Wentzel-Kramers-Brillouin (WKB) approximation, which presumes that $q[i]$ slowly changes across successive slabs (slab i to slab $i \pm 1$). For slowly varying $q[i]$, the WKB approximation states that the change in phase across consecutive slabs is cumulative and the phase retains *memory* from the previous slabs the wave propagated through.

Use of the WKB approximation requires the ionosphere to be a slowly varying stratified medium over the range of the phase integral, and this requirement makes the WKB method a good tool for a class of physical models known as ray tracing, in which a medium is divided into small slabs and Snell's law is solved at each boundary to track the evolving direction

of a wave. The required WKB conditions breaks down, however, for VLF reflections at D region altitudes, because the region within which the VLF wave reflects is now substantially greater than a wavelength. At higher altitudes in the ionosphere or in the magnetosphere, starting around 150 km for VLF waves [*Pappert and Hitney*, 1982] or when X is on the order of 10^3 or 10^4 [*Budden*, 1955a], the WKB method can be used because the electron density is slowly varying compared to VLF wave lengths. The WKB method is also valid for higher frequency propagation. Although the WKB approximation is not valid in the cases we focus on, we describe it nonetheless for completeness.

Using the WKB approximation gives first-term power series accuracy when $\mathrm{d}q/\mathrm{d}z$ is small [*Wait*, 1970, p. 90]. Taking the cumulative phase into account, the more general field expression up to altitude h is

$$F_h(z) = F_0 \exp\left[-jk\left(x\sin\theta + \int_0^h q(z)\mathrm{d}z\right)\right], \tag{2.22}$$

where F_0 is some constant. In an isotropic ionosphere, the reflection coefficient becomes the ratio of the upward and downward traveling waves

$$R = j\exp\left[-2jk\int_0^h q(z)\mathrm{d}z\right], \tag{2.23}$$

and the more complicated anisotropic case has reflection coefficients that are proportional to a factor

$$R \propto j\exp\left\{-2jk\int_0^h \left[q_1(z) - q_2(z)\right]\mathrm{d}z\right\}, \tag{2.24}$$

where $q_1(z)$ refers to the upward traveling wave and $q_2(z)$ is the downward traveling counterpart [*Budden*, 1985, p. 195].

We next turn to a more general method to solve the coupled Ampere's law and Faraday's law equations, known as the full wave method. Unlike the WKB method, the full wave method is valid even when the permittivity has appreciable change over lengths comparable to or less than a wavelength. Thus, the full wave method is a good approach for finding the reflection matrix for VLF waves in the D region. In the WKB method, we used the fact that the ionosphere is horizontally stratified with slabs that have a width of δz. These successive

slabs will be used again to find a full wave solution. We will first find an initial reflection matrix by solving the Booker quartic at some height well above the effective ionospheric reflection height. The waves from the great height are then followed downward toward the ground, while also integrating the change in reflection matrix for each slab. Eventually, an altitude is reached that is approximately free space, and at this point the ionospheric reflection matrix is obtained.

2.2.5 Accounting for Earth's Curvature

Earth's curvature needs to be accounted for in long-distant EIWG propagation. A direct approach, to account for Earth's curvature, is to build off the mathematics for parallel-plate waveguide theory and develop the mathematics for a waveguide that consists of two concentric conducting cylinders of different radiuses. This approach has been used before, but the mathematics become complicated when using a cylindrical coordinate system [*Wait*, 1970, Ch. 14]. A somewhat simpler approach is to assume the Earth is flat and introduce a correction term in the parallel-plate waveguide mathematics.

Figure 2.3 shows a ray (red line) referenced to a spherical surface (left panel) that is then transformed to reference a flat surface (right panel). Earth's curvature causes the distance d between the ray and the spherical boundary to change non-linearly, and the ray for the flat boundary case perfectly mimics this non-linear change in distance. An observer who is standing on either boundary, and is oblivious to whether the surface is spherical or flat, would perceive the ray to curve upward away from the boundary in the exact same way for both boundaries.

Budden and Martin [1962] use this spherical to flat boundary transformation to devise an elegant solution that accounts for Earth's curvature by supposing the Earth is flat and the free space between the Earth and ionosphere is an isotropic medium that approximately satisfies

$$\frac{\mathrm{d}n}{\mathrm{d}z} = \frac{1}{R_e}, \tag{2.25}$$

where R_e is the radius of Earth. Since the radius of the Earth is much larger than the height of the ionosphere, there is some freedom in choosing the height z_0 where the index

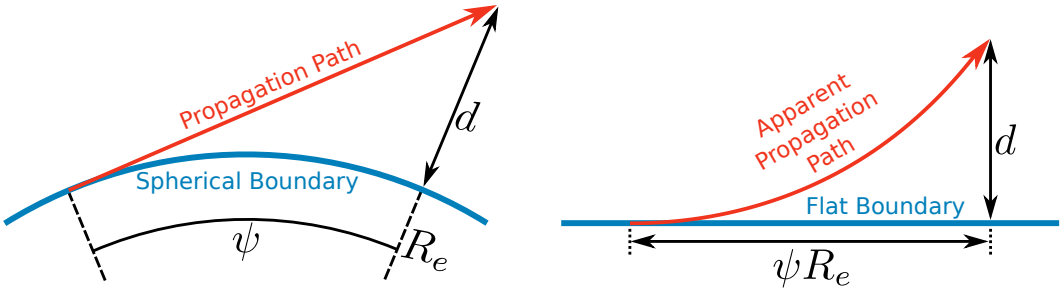

Figure 2.3: Depiction of a ray (red line) propagating in a straight line away from a spherical boundary (left panel) transformed into a curved ray propagating away from a flat boundary (right panel).

of refraction will be unity.

This technique is analogous (but in reverse) to a technique called *effective Earth radius* often used to trace propagation of GHz frequencies through free space. Because the index of refraction of air (at GHz frequencies) is slightly more than 1 at sea level and decreasing with altitude, there is some refraction of any transmitted wave. The most common way to correct this is to presume the Earth's radius is actually some multiple of its true radius, commonly 4/3. This then *unbends* the ray by just the right amount so as to straighten it. In our application, we are doing the opposite, we are flattening the Earth, but then adding artificial refraction to bend each wave in just the right way as to account for the Earth's curvature. In particular, *Budden and Martin* [1962] found that

$$n^2 = 1 - \frac{2(z_0 - z)}{R_e} \tag{2.26}$$

gives a good approximation to (2.25) (see *Richter* [1966] for additional details).

Solving the Booker quartic required the use of the susceptibility matrix $\mathbf{M}$ (2.15). To adjust for Earth's curvature, we simply include the additional index of refraction term into the susceptibility matrix

$$\tilde{\mathbf{M}} = \mathbf{M} - \frac{2(z_0 - z)}{R_e}\mathbf{I} \tag{2.27}$$

[*Morfitt and Shellman*, 1976]. With this adjustment, curvature errors from integration over the susceptibility matrix as a function of altitude will be reduced.

2.2.6 Ionospheric Reflection Matrix

The full wave method requires the coupling of Ampere's Law (1.1) and Faraday's Law (1.2), and we wish to find a relation between the changing fields as a function of height. *Clemmow and Heading* [1954] formulated a method to solve the differential equation with respect to height

$$\mathbf{e}' = -j\mathbf{T}\mathbf{e}, \tag{2.28}$$

where the $'$ is the $\partial/\partial(kz)$ derivative, the field vector is

$$\mathbf{e} = \begin{bmatrix} E_x \\ -E_y \\ \eta_0 H_x \\ \eta_0 H_y \end{bmatrix}, \tag{2.29}$$

and

$$\mathbf{T} = \begin{bmatrix} \frac{-\widetilde{M}_{31}\sin\theta}{1+\widetilde{M}_{33}} & \frac{\widetilde{M}_{32}\sin\theta}{1+\widetilde{M}_{33}} & 0 & \frac{\cos^2\theta+\widetilde{M}_{33}}{1+\widetilde{M}_{33}} \\ 0 & 0 & 1 & 0 \\ \frac{\widetilde{M}_{23}\widetilde{M}_{31}}{1+\widetilde{M}_{33}} - \widetilde{M}_{21} & \cos^2\theta + \widetilde{M}_{22} - \frac{\widetilde{M}_{23}\widetilde{M}_{32}}{1+\widetilde{M}_{33}} & 0 & \frac{\widetilde{M}_{23}\sin\theta}{1+\widetilde{M}_{33}} \\ 1+\widetilde{M}_{11} - \frac{\widetilde{M}_{13}\widetilde{M}_{31}}{1+\widetilde{M}_{33}} & \frac{\widetilde{M}_{32}\widetilde{M}_{13}}{1+\widetilde{M}_{33}} - \widetilde{M}_{12} & 0 & \frac{-\widetilde{M}_{13}\sin\theta}{1+\widetilde{M}_{33}} \end{bmatrix}. \tag{2.30}$$

Notice that variable $\widetilde{M}$ in matrix $\mathbf{T}$ contains the charged species profiles N_α and collision frequency profiles ν_α, which is the exact altitude variation we are concerned with. The field vectors are related to the four solutions to the Booker quartic

$$\begin{aligned} &\text{upward ordinary wave:} && E_x = \eta_0 H_y \cos\theta, \quad \eta_0 H_x = E_y = 0, \\ &\text{upward extraordinary wave:} && \eta_0 H_x = -E_y \cos\theta, \quad \eta_0 H_y = E_x = 0, \\ &\text{downward ordinary wave:} && E_x = -\eta_0 H_y \cos\theta, \quad \eta_0 H_x = E_y = 0, \\ &\text{downward extraordinary wave:} && \eta_0 H_x = E_y \cos\theta, \quad \eta_0 H_y = E_x = 0. \end{aligned} \tag{2.31}$$

At any altitude, the total field is the sum of the upward traveling wave field with its downward traveling counterpart field, this gives the matrix relation

$$\mathbf{e} = \begin{bmatrix} \cos\theta & 0 & -\cos\theta & 0 \\ 0 & -1 & 0 & -1 \\ 0 & -\cos\theta & 0 & \cos\theta \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \tag{2.32}$$

or more compactly

$$\mathbf{e} = \mathbf{L}\mathbf{q}, \tag{2.33}$$

where the elements of $\mathbf{q}$ are the complex amplitudes of the fields [*Barron and Budden*, 1959]. Note that a and b are for upward traveling fields and c and d are for downward traveling fields, so

$$\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} c \\ d \end{bmatrix}, \tag{2.34}$$

and these vectors are related to the reflection matrix through

$$\mathbf{d} = \mathbf{R}\mathbf{u}, \tag{2.35}$$

which is the same as (2.8).

A solution for $\mathbf{R}$ is found by integrating over the small changes in the reflection matrix $\mathbf{R}'$ as a function of altitude, which is related to (2.28). Rearranging (2.33) gives $\mathbf{q} = \mathbf{L}^{-1}\mathbf{e}$, and the derivative is found using (2.28)

$$\begin{aligned} \mathbf{q}' &= -j\mathbf{L}^{-1}\mathbf{T}\mathbf{L}\mathbf{q} \\ &= \frac{-j}{2}\mathbf{S}\mathbf{q}. \end{aligned} \tag{2.36}$$

The solution to $\mathbf{S} = 2\mathbf{L}^{-1}\mathbf{T}\mathbf{L} \in \mathbb{C}^{4\times 4}$ is given by *Budden* [1955a], and is written in a way

that sections the matrix into four distinct $\mathbb{C}^{2\times 2}$ sub-matrices

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}^{11} & \mathbf{S}^{12} \\ \mathbf{S}^{21} & \mathbf{S}^{22} \end{bmatrix}. \tag{2.37}$$

Recognizing that $\mathbf{q}$ contains the upward and downward wave coefficients, the derivative of (2.34) is

$$\begin{aligned} \mathbf{u}' &= \frac{-j}{2}\left[\mathbf{S}^{11}\mathbf{u} + \mathbf{S}^{12}\mathbf{d}\right] \\ \mathbf{d}' &= \frac{-j}{2}\left[\mathbf{S}^{21}\mathbf{u} + \mathbf{S}^{22}\mathbf{d}\right]. \end{aligned} \tag{2.38}$$

Taking the derivative of (2.35) gives the derivative of the reflection matrix we seek

$$\mathbf{d}' = \mathbf{R}\mathbf{u}' + \mathbf{R}'\mathbf{u}. \tag{2.39}$$

Arranging terms and using (2.38) gives the final differential reflection matrix

$$\mathbf{R}' = \frac{-j}{2}\left[\mathbf{S}^{21} + \mathbf{S}^{22}\mathbf{R} - \mathbf{R}\mathbf{S}^{11} - \mathbf{R}\mathbf{S}^{12}\mathbf{R}\right]. \tag{2.40}$$

It should be noted that LWPC integrates over a modified reflection matrix $\mathbf{X}'$, where $\mathbf{X} = (\mathbf{R} + \mathbf{I})/\cos\theta$, for numerical convenience, but we limit our discussion to just $\mathbf{R}'$.

Budden [1955a] uses Runge-Kutta to solve the ordinary differential equation (2.40). Runge-Kutta is an iterative method for numerically integrating ordinary differential equations. When beginning the Runge-Kutta integration, some initial condition must be known. At a high enough altitude, much higher than the effective reflection height, the reflection between two slabs is approximated to be a sharp boundary between two homogeneous plasma mediums. Thus, the method from *Sheddy* [1968], which uses the Booker quartic, is a valid approach for finding the initial conditions of the reflection matrix at this great altitude. The Runge-Kutta method then iterates over (2.40) in small downward δz steps. It is important to choose an integration step size that is small enough to avoid a loss in numerical accuracy, but not too small or else the computation time will be unnecessarily long. A good starting point in choosing the integration step size is $\delta z = \lambda/100$ [*Budden*, 1985, p. 570]. Integration

will continue until an altitude b, that is approximately free space and has plasma effects that are so weak that they are negligible, has been reached. *Budden* [1955b] estimates this minimum altitude to be when $X \approx 1$. Completion of the integration gives the final reflection matrix

$$\mathbf{R}_b(\theta) = \begin{bmatrix} {}_{\parallel}R^b_{\parallel}(\theta) & {}_{\perp}R^b_{\parallel}(\theta) \\ {}_{\parallel}R^b_{\perp}(\theta) & {}_{\perp}R^b_{\perp}(\theta) \end{bmatrix} \tag{2.41}$$

for the anisotropic and vertically inhomogeneous ionosphere at an altitude of b.

2.3 Mode Finding

In (2.6) resonant modes are found using reflection coefficients that are referenced to their respective boundaries, and the $\exp[-j2ak\cos\theta_m]$ term is necessary to relate the two reflecting coefficients together. This exponential term, however, is only valid for parallel boundaries and does not extend to the more complicated geometry of Earth's curvature. From (2.11) the ground reflection matrix $\overline{\mathbf{R}}_g(\theta)$ is referenced to an altitude of zero, and from (2.41) the ionospheric reflection matrix $\mathbf{R}_b(\theta)$ is referenced to altitude b. It is evident that some mechanism is needed to relate the two reflection matrices to each other while accounting for Earth's curvature.

A modification is applied to the free space index of refraction (2.26) in the susceptibility matrix (2.27) to account for Earth's curvature during the integration when finding $\mathbf{R}_b(\theta)$. Using this same concept, an adjusted free space susceptibility matrix is found to be $\widetilde{\mathbf{M}}_f = [1-2(z_0-z)/R_e]\mathbf{I}$. Runge-Kutta integration for free space (using $\widetilde{\mathbf{M}}_f$ instead of $\widetilde{\mathbf{M}}$) is now performed on $\overline{\mathbf{R}}_g(\theta)$ and $\mathbf{R}_b(\theta)$ in an upward manner (towards the ionosphere). Integration is complete when altitude d is reached, giving reflection matrices $\overline{\mathbf{R}}_d(\theta)$ and $\mathbf{R}_d(\theta)$. *Morfitt and Shellman* [1976] modified the integration algorithm in LWPC to a method similar to that described here but more computationally efficient.

With the reflection matrices referenced to the same altitude, the modal equation from (2.6) is expanded to the reflection matrix case $\mathbf{R}_d(\theta)\overline{\mathbf{R}}_d(\theta) = \mathbf{I}$, and the modes are found by solving

$$\det\left[\mathbf{I} - \mathbf{R}_d(\theta)\overline{\mathbf{R}}_d(\theta)\right] = 0, \tag{2.42}$$

where complex values of θ that solve the equation are resonant modes (complex eigenangles) for the EIWG. Eigenangle solutions to the determinant reside in a complex plane, and finding these solutions is a non-trivial task. An algorithm from *Shellman* [1986] is used in LWPC to find the dominant resonant modes (eigenangles with the largest real component) and the ideal reference altitude d.

Qualitative reasoning is used to find the modal function (2.42), but it does not give insight into the excitation or propagation of radio waves within the waveguide. *Budden* [1962] showed a rigorous method for the excitation of fields in the waveguide, and this method has since been adapted for use in LWPC [*Pappert and Ferguson*, 1986]. Budden's method begins by assuming a vertical line of quadrupole sources, of strength Q_U, exist on the $\hat{\mathbf{z}}$ axis centered at the origin. This radiating source is represented by the electric Hertz vector (the Hertz vector is related to the scalar electric potential through $\Phi = -\nabla \mathbf{U}$ and is also often written as $\mathbf{\Pi}$)

$$U_z = \frac{kQ_1}{4\pi} \int_{-j\infty}^{\pi+j\infty} \exp\left[-jk\left(x\cos\phi + z\sin\phi\right)\right]\cos\phi \, \mathrm{d}\phi, \tag{2.43}$$

where ϕ is the angle from the xz-plane relative to the $\hat{\mathbf{x}}$ axis (Budden alternates the use of θ as the angle relative to the bottom or top of the waveguide, but this is confusing so ϕ is used to avoid misunderstandings). A similar expression also exists for the case where the source is a magnetic line quadrupole, which would result in the magnetic Hertz vector V_z, which is the counterpart to U_z, with a strength of Q_V.

Fields from U_z and V_z are polarized in the TM and TE directions, respectively. These two orthogonal polarizations are written as the vector

$$\mathbf{h} = \begin{bmatrix} U_z \\ -V_z/\eta_0 \end{bmatrix} \tag{2.44}$$

and are related to the magnetic and electric fields in the transverse ($\hat{\mathbf{y}}$) direction by

$$\begin{bmatrix} \eta_0 H_y \\ E_y \end{bmatrix} = -j\omega\eta_0 \frac{\partial \mathbf{h}}{\partial x}. \tag{2.45}$$

Using a technique similar to (2.35), we can define a set of upward $\mathbf{h}_u$ and downward $\mathbf{h}_d$ traveling waves and relate them through a reflection matrix

$$\mathbf{h}_d = \mathbf{R}\mathbf{h}_u. \tag{2.46}$$

Budden then shows that the electric and magnetic Hertz vectors are related to the ground and ionosphere reflection matrices, relative to height d, through

$$\mathbf{h} = \frac{k}{4\pi} \int_{-\pi/2-j\infty}^{\pi/2+j\infty} \exp\left[-jk\left(x\cos\phi + z\sin\phi\right)\right] (\mathbf{I} + \mathbf{R}_d)\mathbf{W}(\mathbf{I} + \overline{\mathbf{R}}_d)\mathbf{Q}\cos\phi \, \mathrm{d}\phi, \tag{2.47}$$

with $\mathbf{W} = [\mathbf{I} - \overline{\mathbf{R}}_d\mathbf{R}_d]^{-1}$ and $\mathbf{Q} = [Q_U, Q_V]^T$. Notice that the limits of integration for the contour integral have been distorted from $-j\infty$ through $\pi + \infty$ in (2.43) to $-\pi/2 - j\infty$ through $\pi + j\infty$, doing so makes the integral symmetric about $\phi = 0$. Integration over the contour reveals the existence of singularities, most importantly poles. These poles reside at locations where $\det\left[\mathbf{W}^{-1}\right] = 0$ is satisfied, which is equivalent to the modal function in (2.42). Each pole inside the contour represents a single resonant mode, and the total Hertz vectors are found by adding up the individual residue contributions from each pole (note that each residue is multiplied by a factor of $j2\pi$, this comes from the residue theorem)

$$\mathbf{h} = \frac{jk}{2} \sum_m \mathbf{\Lambda}_m \mathbf{Q} e^{-jk(x\cos\phi_m + z\sin\phi_m)}, \tag{2.48}$$

where

$$\mathbf{\Lambda_m} = \frac{[\mathbf{I} + \mathbf{R}_d(\phi_m)]\mathbf{X}_m[\mathbf{I} + \overline{\mathbf{R}}_d(\phi_m)]}{\frac{\partial\Delta}{\partial(\sin\phi)}\Big|_{\phi=\phi_m}} \tag{2.49}$$

is referred to as the height gain function, $\Delta = \det\left[\mathbf{W^{-1}}\right]$, and $\mathbf{X}_m = \lim_{\phi\to\phi_m} \mathbf{W}\Delta$. Combining (2.45) with (2.48) gives an expression for the total modal contributions to the transverse magnetic and electric fields in the waveguide.

Pappert and Shockey [1971] extended the quadrupole source method from *Budden* [1962] to a more general line source that may be oriented in any direction and at any altitude (within the waveguide). Likewise, a receiver may not always be located on the ground and fields other than H_y and E_y may be desired. *Morfitt and Shellman* [1976] and *Pappert and*

Ferguson [1986] give an equation for an arbitrary field (electric or magnetic field in any direction), with lookup tables for different sources and fields, in the functional form

$$F(x) = Q_F \left[\sin\left(\frac{x}{R_e}\right)\right]^{-1/2} \sum_m \Lambda_{tm}(z_t)\Lambda_{rm}(z_r)e^{-jkx\sin\theta_m}, \tag{2.50}$$

where Q_F is a field dependent constant, Λ_{tn} and Λ_{rn} are the transmitter and receiver height gain functions, and z_t and z_r are the transmitter and receiver heights. Budden's method assumes a flat waveguide for mathematical convenience, therefore the correction term $1/\sqrt{\sin(x/R_e)}$ is included to account for Earth's curvature [*Wait*, 1970, p. 161]. In practice, waves traveling along the opposite side of Earth produce significant modal interference near the antipode, but this method does not account for these opposite side traveling waves and does not hold well for distances close to half the circumference of Earth.

Equation (2.50) reveals an important, but not so obvious, property about the coupling of the QTM and QTE polarizations. The only term in (2.50) that depends on x is $\exp(-jkx\sin\theta_m)$, which is the term that dictates the attenuation

$$\alpha_m = -k\Im\{\sin\theta_m\}, \tag{2.51}$$

and phase velocity

$$v_{pm} = \frac{c}{\Re\{\sin\theta_m\}} \tag{2.52}$$

of the field [*Pappert et al.*, 1967]. Eigenangles for lower order modes have larger real values and smaller imaginary values, so smaller mode numbers have lower attenuation and slower phase velocity. Therefore, for a given mode number, the electric and magnetic fields both have the same attenuation rate and phase velocity, which is not generally true for isotropic waveguide propagation. This interesting effect occurs in the anisotropic case because the off diagonal terms in the ionospheric reflection matrix $\mathbf{R}_d$ are the result of coupling between the TM and TE polarizations, shown in (2.9) and (2.10). Thus, the coupling of the polarizations forces the QTM and QTE to attenuate at the same rate and travel with the same phase velocity.

The work in this thesis is concerned with VLF antenna sources that have a vertical

electric dipole moment M_e and receivers that measure magnetic flux density in the horizontal plane. *Cummer et al.* [1998] gives the field expression for the QTM flux density at height z_r as

$$B_y(x) = -\mu_0 M_e e^{j\pi/4} \sqrt{\frac{2k}{\pi R_e \sin(x/R_e)}} \sum_m \Lambda_{tm}(z_t)\Lambda_{rm}(z_r) e^{-jkx\sin\theta_m}. \tag{2.53}$$

2.4 Mode Conversion

LWPC also has the capability to connect two different waveguide segments, with different ionospheric and ground parameters through the method known as mode conversion. Figure 2.4 gives an example for these different segments, where each segment has its own unique h', β, background magnetic field vector, ground conductivity, and ground relative permittivity. Parameters in the first segment, where the transmitter is, are marked with a subscript 1. If we apply the parameters in segment 1 to the mode finding method previously discussed, then we know B_y along the ground through the first segment. Our knowledge of B_y ends at the first segment because the mode finding method assumes horizontal homogeneity, which segment 2 violates.

We know how to use the mode finding method to find the modes in each segment, but we do not know the conversion mechanism for transferring radio wave power between segment 1 and segment 2. Mode conversion solves this transfer problem by using the known fields at the boundary point of the first segment to determine how much power from each mode in segment 1 is transferred into each individual mode in segment 2. The process is then repeated for the transfer of power between segments 2 and 3, and we then find B_y in segment 3 at the location of our receiver.

Mode conversion was first postulated by *Crombie* [1964] to explain quick phase steps and amplitude fading of recorded VLF signals during day-to-night and night-to-day transitions, known as the day/night terminator. His theory focused on a simplified case where two modes existed in the night side of the waveguide and a single mode existed in the day side. Mode conversion at the night-to-day terminator converts some power from the two nighttime waves into the single daytime wave, and the rest of the unconverted power is lost in the waveguide

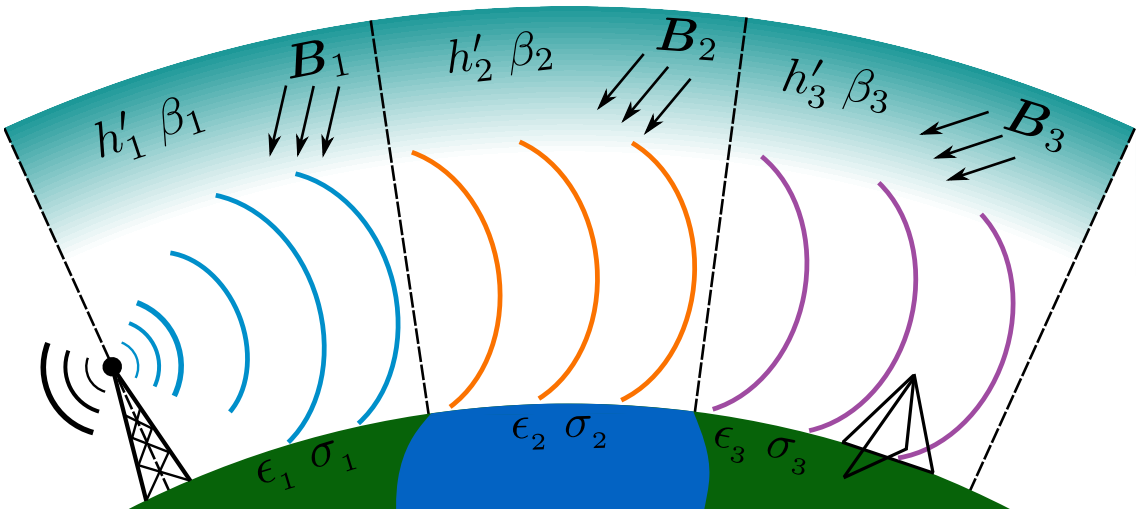

Figure 2.4: LWPC partitions the waveguide into multiple homogeneous segments defined by the user. Each segment is defined by: h', β, conductivity σ, relative permittivity ϵ_r, segment length, and the three-dimensional vector for Earth's background magnetic field.

as evanescent waves. Crombie's theory was then validated by *Walker* [1965]. *Wait* [1968a,b], along with *Galejs* [1971], created a more general mode conversion method where an arbitrary number of modes from one segment would then excite an arbitrary number of modes in the next segment. Contributions to LWPC's mode conversion algorithm comes from *Pappert and Snyder* [1972], *Pappert and Morfitt* [1975], and *Ferguson and Snyder* [1980]. Two mode conversion options exist for LWPC, *Full MC* and *Fast MC*. Much of the details for the mode conversion method involve mathematical endeavors that do not give significant insight into the physics of waveguide propagation. Here, we will briefly discuss some of the fundamental concepts for mode conversion. *Pappert and Snyder* [1972], *Pappert and Morfitt* [1975], and *Ferguson and Snyder* [1980] are suggested readings for more details.

A limitation of LWPC's mode conversion method is that it does not account for backwards scattering from horizontal inhomogeneities. Consider two waveguide regions, regions I and II, that are connected by some scattering discontinuity. The forward traveling modes in region I are related to the forward traveling modes in region II through matrix **A**. There will also be some scattering from the discontinuity that will produce backwards traveling modes in region I, these backwards traveling modes are related through matrix **B**. Finally, there may exist some backwards propagating modes in region II, related through matrix

C, that were produced by some scattering event further down the waveguide. The forward mode m propagating fields $\overrightarrow{\boldsymbol{\alpha}}^{\mathrm{I}}_m$, in region I, are related to the other fields by

$$\overrightarrow{\boldsymbol{\alpha}}^{\mathrm{I}}_m(z) + \sum_i \mathbf{B}_{mi}\ \overleftarrow{\boldsymbol{\alpha}}^{\mathrm{I}}_i(z) = \sum_\ell \mathbf{A}_{m\ell}\ \overrightarrow{\boldsymbol{\alpha}}^{\mathrm{II}}_\ell(z) + \sum_n \mathbf{C}_{mn}\ \overleftarrow{\boldsymbol{\alpha}}^{\mathrm{II}}_n(z). \tag{2.54}$$

LWPC does not account for these backward propagating modes, so the matrices **B** and **C** become zero. Thus, only the forward propagating modes are left

$$\overrightarrow{\boldsymbol{\alpha}}^{\mathrm{I}}_m(z) = \sum_\ell \mathbf{A}_{m\ell}\ \overrightarrow{\boldsymbol{\alpha}}^{\mathrm{II}}_\ell(z). \tag{2.55}$$

This is important to recognize since some D region perturbations produce appreciable backscattering [*Dowden et al.*, 1996; *Marshall et al.*, 2006; *Mika and Haldoupis*, 2008].

LWPC's field calculation for a homogeneous waveguide (horizontally homogeneous for the ionosphere), given in functional form in (2.50), is done by summing the fields for each mode. In the multi-segmented case, the expected field expression for a single mode m in segment p is

$$F_m(x) = \frac{Q_F}{\sqrt{\sin\left(\frac{x}{R_e}\right)}} \sum_i a^p_{im} \Lambda_{tm}(z_t) \Lambda_{rm}(z_r) \Phi^p_{im}\ \exp\left\{-jk\left[x_2 \sin\theta^1_m + (x - x_p)\sin\theta^p_i\right]\right\}\ , \tag{2.56}$$

where $a^p_{im} \in \mathbb{C}$ is a conversion coefficient for the ith mode in segment $p-1$ to the mth mode in segment p, x_p is the distance from the transmitter to the beginning of segment p, and $\Phi^p_{im} = \sin\theta^p_i / \sin\theta^1_m$ when F is B_y or E_z and is unity otherwise [*Pappert and Ferguson*, 1986]. Equation (2.56) is expressed in a form similar to (2.50) by simply summing the modes

$$F(x) = \sum_m F_m(x). \tag{2.57}$$

The major difference between (2.50) and (2.57) is the use of the mode conversion coefficient $a^p_{im}$to relate all modes from segment $p-1$ to mode m in segment p; it is this set of coefficients that the mode conversion method determines.

Fields in the boundary plane (yz-plane) of segment number p, for mode i, are written

as

$$\mathbf{g}_i^p(z) = \begin{bmatrix} E_{yi}^p(z) \\ E_{zi}^p(z) \\ \eta_0 H_{yi}^p(z) \\ \eta_0 H_{zi}^p(z) \end{bmatrix}, \tag{2.58}$$

where the fields are a function of altitude and are invariant in the $\hat{\mathbf{y}}$ direction. These fields are often referred to as height gain functions. In general, it is possible to define another set of height gain functions $\mathbf{f}_i^p(z)$ orthogonal to $\mathbf{g}_m^p(z)$, except when $i = m$,

$$\delta_{im} = \int_{-\infty}^{\infty} [\mathbf{f}_i^p(z)]^\dagger \, \mathbf{g}_m^p(z) \, \mathrm{d}z, \qquad \delta_{im} = \begin{cases} 1 & i = m \\ 0 & i \neq m \end{cases}, \tag{2.59}$$

where † is the adjoint (conjugate transpose) operator. This orthogonality is used to relate mode m in segment $p-1$ and mode i in segment p

$$I_{im}^{p,p-1} = \int_{-\infty}^{\infty} [\mathbf{f}_i^p(z)]^\dagger \, \mathbf{g}_m^{p-1}(z) \, \mathrm{d}z. \tag{2.60}$$

Pappert and Snyder [1972] show that conversion coefficients in segment p are related to the conversion coefficients in segment $p-1$ through

$$\begin{aligned} a_{im}^p &= \sum_i a_{i\ell}^p \; I_{m\ell}^{p,p} \\ &= \sum_\ell a_{\ell m}^{p-1} \; I_{i\ell}^{p,p-1} \exp[-jk(x_p - x_{p-1}) \sin\theta_\ell^{p-1}], \end{aligned} \tag{2.61}$$

where $a_{im}^1 = \exp\left(jkx_2 \sin\theta_i^1\right)$ for the special case when $p = 1$. This relation for the current segment's coefficient to rely on the previous segment's coefficient is similar to the simple scattering example in (2.55).

The *Full MC* method from *Pappert and Snyder* [1972] uses a full wave solution to find $\mathbf{f}_i^p(z)$ and $\mathbf{g}_i^p(z)$. Alternatively, the *Fast MC* method from *Pappert and Morfitt* [1975], which was later modified by [*Ferguson*, 1980], does not use a full wave solution and does not require the orthogonality between $\mathbf{f}_i^p(z)$ and $\mathbf{g}_i^p(z)$ in (2.59). Instead, the *Fast MC* method restricts

the integral of the height gain functions

$$I_{im}^{p,p-1} = \left(\sin\theta_i^p + \sin\theta_m^{p-1}\right) \int_0^b \left[E_{yi}^p \, E_{ym}^{p-1} + \eta_0^2 \, H_{yi}^p \, H_{ym}^{p-1}\right] \, \mathrm{d}z \tag{2.62}$$

to be from the ground to the bottom of the ionosphere, and the fields are solved for using a variation of the non-orthogonal basis expansion from *Wait* [1968a]. *Pappert and Ferguson* [1986] reported good agreement between the *Full MC* and *Fast MC* methods, therefore similar results should be expected with LWPC regardless of the mode conversion method used.

2.5 LWPC Examples

As the diversity of references throughout this chapter suggests, numerous researchers and organizations have contributed to the development of LWPC. The version of LWPC used in this work is known as version 3.0 and is a slight modification of version 2.0 from *Ferguson* [1998] at the US Space and Naval Warfare Systems Command (SPAWAR). The code is almost exclusively written in FORTRAN, and even has some comments that are reminiscent of the code's early development with punch cards. The following are a few examples for the output of LWPC V3.0.

The first example, shown in Figure 2.5, is a simple homogeneous LWPC run with no mode conversion, meaning all parameters are constant throughout the waveguide except for the vertical electron density profile. NAA's transmitter profile, a vertical electric field antenna at 24 kHz with a power of 1.2 MW, is used. Earth's ground is set to have electrical properties similar to salt water, which is a good radio reflector, with $\sigma = 4$ and $\epsilon_r = 81$, and a background magnetic field 45° to zenith and the plane of propagation, with a strength of 35 μT (note the input to LWPC is in Gauss and not μT), is used. The electron density profile is defined using the *Wait and Spies* [1964] two-parameter model (1.22) with the values $\beta = 0.4$ and $h' = 70$. Amplitude of the TM signal is shown in the top panel and an adjusted phase of the signal is shown in the bottom panel. The speed of light phase (kx) has been removed from the TM signal to add clarity in the display of phase changes as a function of distance.

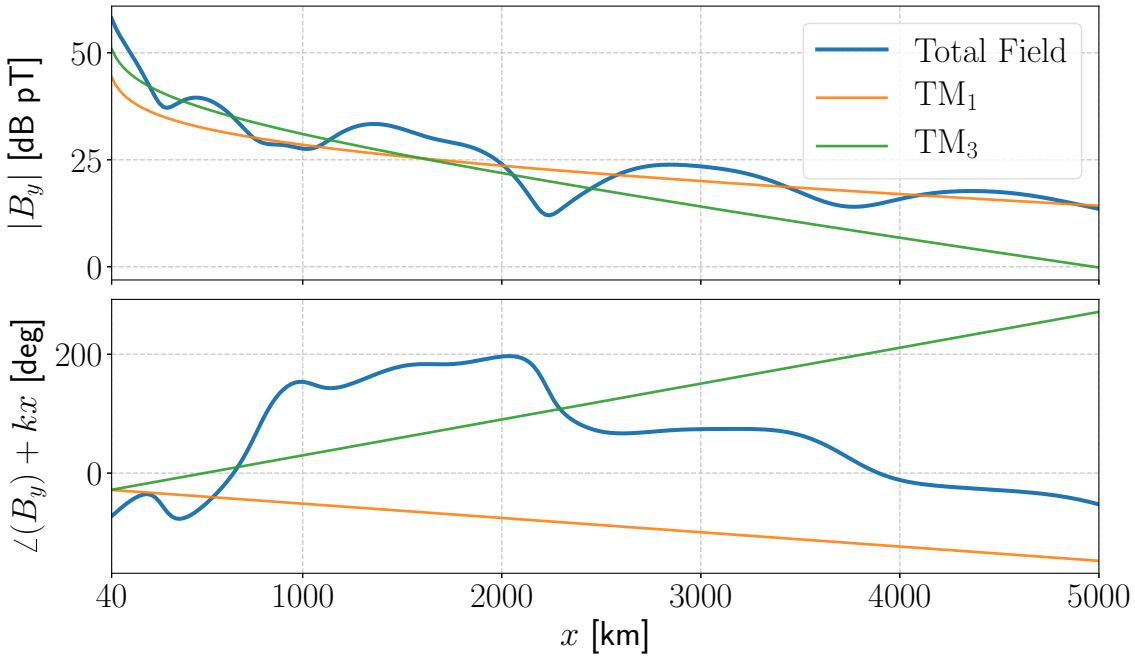

Figure 2.5: Total amplitude and phase from example LWPC run with $\beta = 0.4$ and $h' = 70$. The two strongest modes, TM_1 and TM_3, are also shown.

Mode	v_p/c	α [rad/Mm]	$\lvert\Upsilon\rvert$ [pT]	$\angle\Upsilon$ [deg]
TM_1	0.9976	0.221	13.152	-78.413
TM_2	0.9992	0.483	0.070	3.180
TM_3	1.0060	0.709	28.396	-87.365
TM_4	1.0134	1.313	0.090	19.983
TM_5	1.0251	1.796	24.544	-87.484
TM_6	1.0374	2.761	0.197	23.051

Table 2.1: Example daytime QTM mode values for LWPC run with $\beta = 0.4$ and $h' = 70$.

A few noticeable features are the peaks and nulls in the amplitude signal along with sharp phase changes associated with amplitude nulls. Nulls in the amplitude curve are caused by modal interference (modes with similar amplitudes are close to 180° out of phase and closely cancel with each other). The most noticeable null occurs around 2.2 Mm, where the TM_1 and TM_3 modes have similar amplitudes and are 180° out of phase, relative to each other. The two TM modes shown in the figure also exhibit a non-linear decay when x is small and then approach a linear decay as x becomes larger. This non-linear decay comes from the $[\sin(x/R_e)]^{-1/2}$ curvature adjustment term in (2.50).

The total field in Figure 2.5 comes from the superposition of individual modes in (2.50).

Mode	v_p/c	α [rad/Mm]	$\lvert\Upsilon\rvert$ [pT]	$\angle\Upsilon$ [deg]
TM_1	0.9959	0.106	4.279	-82.689
TM_2	0.9964	0.302	1.356	-67.380
TM_3	1.0022	0.246	25.214	-88.324
TM_4	1.0065	0.563	0.739	-14.835
TM_5	1.0150	0.619	22.289	-90.705
TM_6	1.0226	1.122	1.194	-12.328

Table 2.2: Example nighttime QTM mode values for LWPC run with $\beta = 0.5$ and $h' = 80$.

Table 2.1 gives the necessary components needed to define each individual mode. Phase velocity v_p (2.52), normalized by the free space speed of light, and attenuation α (2.51) are direct results of the eigenvalues found from solving the determinant in (2.42). A field strength parameter $\Upsilon = Q_{\mathrm{TM}}\Lambda_{tm}(z_t)\Lambda_{rm}(z_t)$ is used to account for the excitation of each mode m in waveguide and the mode dependent height gain function of the receiver.

A few general trends are exhibited within the table; phase velocity and attenuation both increase for higher order modes, and magnitude of the field strength parameter is much stronger for odd order modes. From the simple parallel-plate waveguide example in Figure 2.1, it is expected that attenuation increases for higher order modes, since higher order mode numbers have smaller reflection angle, and thus reflect a larger number of time off the lossy ionosphere boundary. Likewise, smaller reflection angles mean the wavelength in the $\hat{\mathbf{x}}$ direction λ_x becomes larger, and so the phase velocity in the $\hat{\mathbf{x}}$ direction increases. Different modes couple more efficiently into the EIWG depending on the geometry of the transmitting antenna and the electrical properties of the waveguide. For this particular daytime example, with a transmitter that a has a vertical electric dipole moment, it happens to be that the odd numbered modes couple into the waveguide much more efficiently than the even number modes. In Figure 2.5, the TM_3 mode is dominate up until ~1.6 Mm and then the TM_1 mode dominates. Strength in the TM_3 mode is seen for short distances because it has a strong Υ value, but the TM_3 mode then attenuates more quickly than the TM_1 mode because it has a larger α value.

A similar LWPC example is given in Table 2.2, except now the ionospheric electron density is adjusted to a nighttime profile, with $\beta = 0.5$ and $h' = 80$. The nighttime ionosphere is at a greater altitude than the daytime ionosphere and has a sharper electron

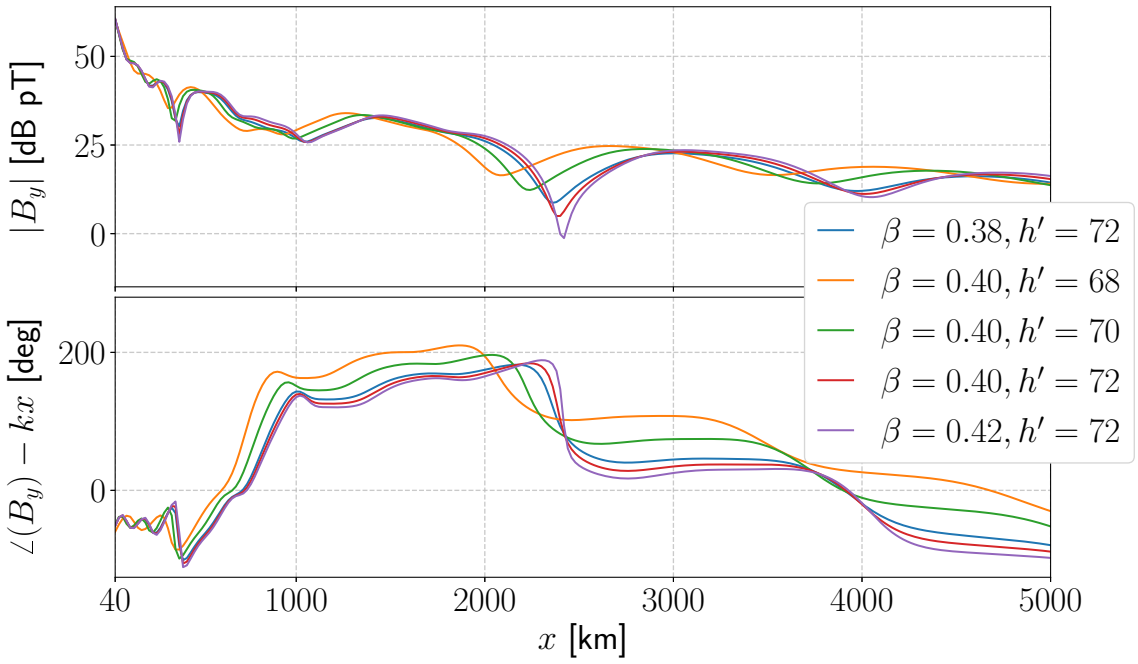

Figure 2.6: Example LWPC runs with various values for h' and β.

density profile. Therefore, the reflection angles are larger in the nighttime, giving slower phase velocities, and the attenuation of the modes is less, since the wave reflects more efficiently off the steep electron density profile. It is also seen that the even number modes have greater $|\Upsilon|$ values in the nighttime than in the daytime. This leads to stronger modal interference patterns during nighttime propagation.

A final example, shown in Figure 2.6, is given to compare small changes in daytime waveguide parameters. The same values used in Figure 2.5 are used here, except the waveguide parameters are varied. The interesting feature in this example is the deep null between the 2 Mm to 2.4 Mm range. By adjusting the waveguide parameters, the intensity of the null is increased or decreased, and the change in phase is made smoother or sharper.

2.6 VLF Remote Sensing History

Narrowband VLF transmitters and lightning are the two most common VLF sources used in D region remote sensing applications. Using narrowband transmitters for remote sensing is advantageous because the transmitted frequency and power is well known, making it easy

to characterize the excitation of a VLF wave in the waveguide. However, these transmitters are sparsely populated around the world and operate over a small bandwidth (often 100 Hz) giving low spectral and spatial resolution. Interestingly, lightning complements the sparse population of narrowband transmitters quite well. Lightning strokes occur over much of the world and give a spatially dense set of VLF emissions. Unfortunately, individual lightning strokes are not well characterized, so the expected received signal from an individual sferic is not well known [*Said et al.*, 2010].

Inferring the h' and β waveguide parameters through VLF remote sensing has traditionally involved an inverse modeling approach. Typical VLF waveguide propagation models use input parameters that define the electrical properties of Earth's ground and the D region along the path of propagation, and the output of the model is often the amplitude and phase of the electric or magnetic field at some point along the waveguide. The inverse modeling approach uses a VLF propagation model and the known amplitude and/or phase of the observed signal to determine h' and β along the path.

Two common methods exist for inverse modeling with narrowband transmitters. The first method consists of a stationary transmitter and a mobile receiver, often mounted to an aircraft [*Wait and Spies*, 1964; *Bickel et al.*, 1970; *Ferguson*, 1980; *Pappert and Hitney*, 1988]. The mobile receiver records the amplitude and phase of the signal while moving along the GCP towards or away from the transmitter. This type of transmitter observation records spatial variations (peaks and nulls) in amplitude and phase due to multi-mode interference. Various combinations of h' and β are then used in a propagation model, and the set of waveguide parameters that gives the best fit to the observed data, with special attention placed towards the peaks and nulls, is then determined. Care must be taken when using the mobile receiver method because variations in the D region may occur during the recording period.

The second, and most common method consists of one or more stationary receivers recording transmissions from stationary transmitters [*Thomson*, 1993; *Clilverd et al.*, 1999; *McRae and Thomson*, 2000, 2004; *Thomson and Clilverd*, 2001; *Bainbridge and Inan*, 2003; *Thomson et al.*, 2011, 2017; *Hayes et al.*, 2017]. Similar to the first method, various combinations of h' and β are used in a propagation model, often LWPC, and the output is

compared with the observed data. Stationary receivers are advantageous for long recording campaigns and observing temporal variations in the D region. Past works have focused on using only one or two receivers due to the high cost of operating a receiver and the increased computational burden every time another receiver is added to the inverse model method.

Broadband VLF radiation from lightning is also used to infer the h' and β waveguide parameters. *Cummer et al.* [1998] and *Han and Cummer* [2010a,b] recorded sferic amplitude with VLF receivers and used an inverse modeling method to infer aggregate h' and β values over the entire sferic spectra. *McCormick et al.* [2018] recently developed an inverse modeling technique with LWPC and used both sferic amplitude and sferic phase to show some frequency dependence in the waveguide parameters.

CHAPTER 3
NARROWBAND VLF FUNDAMENTALS

The terms broadband and narrowband are somewhat ambiguous in general engineering practice. Broadband data means that the data contains a wide spectrum of information, while narrowband data contains a small spectrum of information. The VLF community has adopted a more specific definition for these two terms. In VLF literature, broadband data is the raw, or unaltered, data from a wideband VLF receiver, sampled at baseband. Broadband data from the Stanford AWESOME receiver in *Cohen et al.* [2010a] is defined by the limitations of the receiver, where the bandwidth is 0.3−47 kHz and the sampling rate is 100 kHz. Broadband data from the Georgia Tech LF receiver is again defined by the limitations of the receiver, which has a larger bandwidth (0.5−470 kHz) and a higher sampling rate (1 MHz) [*Cohen et al.*, 2018a]. Storing and working with continuous 1 MHz broadband data is often unwieldy because of the memory requirements (one week of 1 MHz sampled data is ∼1.2 TB). When performing pure VLF research, most of the spectrum in 1 MHz broadband data is irrelevant, and the use of 100 kHz broadband data is better suited for analysis to ease the memory and computational burden. While the Georgia Tech LF receiver does sample at 1 MHz, it also has the capability to downsample the 1 MHz broadband data to a more manageable 100 kHz broadband data set with a bandwidth of 0.5−47 kHz (similar to that in the Stanford receiver) and save the 100 kHz broadband data to memory. This thesis is concerned with signals below the 47 kHz range, so 100 kHz broadband data is sufficient for all the work presented here.

The VLF community loosely uses the term narrowband data to mean small spectrum transmissions in the VLF/LF range. Any small spectrum type of signal may be considered a narrowband signal, but most VLF literature focuses on continuous-wave (CW) or minimum-shift keying (MSK) types of transmissions. CW transmissions are monochromatic so the bandwidth, in theory, is infinitely small. MSK signals generated in the VLF range often, but not always, consist of two tones spaced 100 Hz apart and have side bands that are

contained well within 500 Hz of either side of the center frequency. Therefore, we define a narrowband signal in this thesis as a VLF/LF transmitted signal with a bandwidth less than 1 kHz.

There are two methods for extracting narrowband data with a VLF receiver. The first method requires dedicated hardware that is specifically designed to extract narrowband information for the given type of narrowband transmission. This type of method is often used in industry and commercial products. For example, radio-controlled clocks in the US use dedicated hardware to monitor a 60 kHz CW amplitude modulated signal from the WWVB transmitter, operated by the National Institute of Standards and Technology (NIST), in Fort Collins, Colorado. The second method, and most often used for VLF research, is the extraction of narrowband data from broadband data in either real-time or post-processing. This method is preferred for VLF research because broadband data gives more versatility to the researcher, where a single broadband data set contains broadband events, such as whistlers and sferics, and multiple narrowband signals.

Even though narrowband data has been used for decades in VLF remote sensing applications, little has been published on methods for extracting macroscopic features, such as amplitude and carrier phase, from narrowband VLF signals. *Thomson* [1975, 1981] used a cross-correlation method to estimate MSK signal group delay for whistler mode wave studies. *Shafer* [1994] developed an MSK demodulation method as a means to remote sense transient ionospheric disturbances. *Said* [2009] used a coherent MSK removal technique to reduce MSK transmitter noise in sferic observations. *Mitchell* [2015] created a spread spectrum technique to better estimate the transmitted signal and its response to the EIWG in a multimode environment. *Gross et al.* [2018] extended the coherent MSK demodulation technique to synchronize the demodulation between two co-located antennas for polarization ellipse analysis. This chapter aims to give a clear and detailed discussion of the *Gross et al.* [2018] MSK feature extraction method and discuss how these features may be interpreted with regard to VLF remote sensing.

3.1 Minimum-Shift Keying

3.1.1 MSK Definition

MSK is a modulation scheme that is a special case of Continuous Phase Frequency-Shift Keying (CPFSK), in which the transmitter alternates between two frequencies and has continuous phase (no sharp phase changes between symbol transitions). See *Pasupathy* [1979] for a good introduction on MSK signals. By design, MSK uses two frequencies that are close to the center frequency while still maintaining continuous phase and orthogonality between the upper and lower frequencies. This means that MSK modulation is spectrally efficient because the tones are close together and the sidebands will be small since no sharp phase changes occur.

An ideal MSK signal $s(t)$ with a radial center frequency of w_c contains a stream of bits b_k with values ± 1. For most MSK VLF transmitters, this stream of bits is independent and identically distributed (IID), meaning that each bit is independent of the bits preceding (and following) it. The MSK modulation scheme creates a time-varying signal defined as

$$s(t) = \sum_{k=0}^{\infty} \cos\left[w_c t + \frac{b_k \pi t}{2T} + \theta_k\right] p_T(t - kT), \tag{3.1}$$

where T is the symbol period, θ_k is a phase adjustment to maintain phase continuity, defined by

$$\theta_k = \begin{cases} \frac{-b_k \pi k}{2} + \frac{\pi}{2} \sum_{j=0}^{k-1} b_j, & k > 0 \\ 0, & k = 0 \end{cases}, \tag{3.2}$$

and $p_T(t)$ is a function used to isolate each symbol and minimize inter-symbol interference (ISI) [*Carlson et al.*, 2002, p. 622]. For this work a simple rectangle function will be used as the pulse function

$$p_T(t) = u(t) - u(t - kT), \tag{3.3}$$

where $u(t)$ is the unit step function. Other types of pulse functions exist for MSK, such as the classic Gaussian pulse function which is given the special name Gaussian minimum shift-keying (GMSK) [*Morelli and Mengali*, 1999].

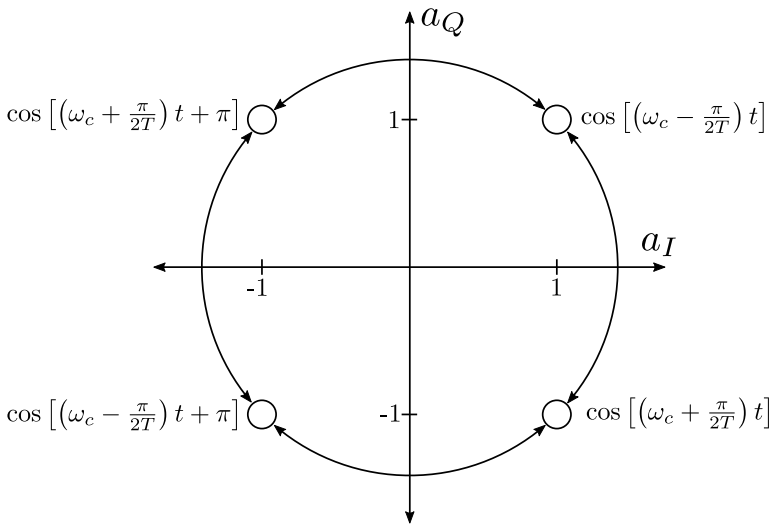

Figure 3.1: In-phase and quadrature symbol mapping.

Inspection of (3.1) combined with (3.2) reveals that $s(t)$ can take on four unique states (tones), which consists of either an upper and lower frequency $\omega_c \pm \pi/(2T)$ and a phase of either 0 or π. Whether $s(t)$ takes on an upper or lower frequency is determined by the value of b_k, meaning the communicated bit stream of 1's and 0's is simply determined by the frequency shift, or lack of a shift, of $s(t)$ every T seconds. The four states have been mapped onto a symbol mapping chart, shown in Figure 3.1, where a_I and a_Q can take on a value of ± 1. A symbol change will occur when $b_k \neq b_{k+1}$, which means $s(t)$ will transition from its current symbol state to an adjacent symbol on the chart, but no symbol change will occur when $b_k = b_{k+1}$. Two choices exist when transitioning to an adjacent symbol, but only one of the choices will allow the signal $s(t)$ to maintain phase continuity during the transition. Equation (3.2) shows that $\theta_k = \theta_{k+1}$ when $k+1$ is an even value, regardless the values of b_k and b_{k+1}, which implies a_I does not change ($a_I[k] = a_I[k+1]$ when $k+1$ is even). Conversely, a_Q does not change when $k+1$ is odd because $s(t)$ must have a phase adjustment of π to maintain phase continuity ($a_Q[k] = a_Q[k+1]$ when $k+1$ is odd).

Each of the four states for $s(t)$ are encoded using a_I and a_Q, meaning the time series

in-phase and quadrature encoding is

$$\begin{aligned} a_I(t) &= \sum_k a_I[k]\; p_{2T}(t - T - 2Tk) \\ a_Q(t) &= \sum_k a_Q[k]\; p_{2T}(t - 2Tk), \end{aligned} \tag{3.4}$$

where $a_I[k]$ and $a_Q[k]$ are gated by

$$p_{2T} = u(t) - u(t - 2kT), \tag{3.5}$$

which is a rectangle function similar to p_T but with a period of $2T$. Combining equations (3.4) and (3.1) gives the easier to interpret MSK time series function

$$s(t) = \cos\left[\omega_c t + b(t)\frac{\pi t}{2T} + \theta(t)\right], \tag{3.6}$$

given

$$\begin{aligned} b(t) &= \begin{cases} +1, & a_I(t) \neq a_Q(t) \\ -1, & a_I(t) = a_Q(t) \end{cases} \\ \theta(t) &= \begin{cases} 0, & a_I(t) = 1 \\ \pi, & a_I(t) = -1 \end{cases} . \end{aligned} \tag{3.7}$$

Equation (3.6) shows that two time series variables, $b(t)$ and $\theta(t)$, are superimposed with the carrier frequency. These two variables are combined into a single time series term known, as a phase trellis

$$\phi_{\text{MSK}}(t) \equiv b(t)\frac{\pi t}{2T} + \theta(t), \tag{3.8}$$

that encapsulates the modulated information.

Equation (3.6) is useful when discussing a simple MSK signal, but it is obviously incomplete because it does not account for time delay of the carrier signal or the phase trellis. We now introduce the term $\tau_0(t)$ (a proxy for phase velocity) to account for any delay of the carrier signal and the term $\tau_1(t)$ (a proxy for group velocity) to adjust the delay of the

symbol transition time

$$x(t) = \cos\left[\omega_c(t - \tau_0(t)) + \phi_{\text{MSK}}(t - \tau_1(t))\right]. \tag{3.9}$$

Any change in τ_0 will phase shift the carrier signal, so we define carrier phase as

$$\phi_0(t) \equiv -\omega_c \tau_0(t). \tag{3.10}$$

The MSK signal has a clock (symbol) frequency of $1/T$, and any delay or advancement of the bit transition time can be thought of as a phase adjustment to the clock frequency. Using (3.8) to expand $\phi_{\text{MSK}}(t - \tau_1)$ shows that clock phase is defined as

$$\phi_1(t) \equiv \frac{-\pi\tau_1(t)}{2T}. \tag{3.11}$$

Now that the carrier phase and clock phase have been defined, the final form of the MSK signal is found by substituting (3.10) and (3.11) into (3.9)

$$\begin{aligned} x(t) &= \cos\left[\omega_c t + \phi_0(t) + \phi_{\text{MSK}}(t - \tau_1(t))\right] \\ &= \cos\left[\omega_c t + \phi_0(t) + b(t - \tau_1(t))\left(\frac{\pi t}{2T} + \phi_1(t)\right) + \theta(t - \tau_1(t))\right]. \end{aligned} \tag{3.12}$$

3.1.2 MSK Demodulation

The first step in demodulating an MSK signal is to isolate the narrowband signal $A(t)x(t)$ from the broadband data $X(t)$. Variable $A(t)$ represents the amplitude of the narrowband signal ($A(t)$ was left out of the previous discussion for clarity but is now being introduced). Isolation is performed by mixing down the narrowband signal to baseband and applying a lowpass filter

$$\begin{aligned} z(t) &= \text{LPF}\{2X(t)e^{-j\omega_c t}\} \\ &= A(t)\exp\left[j\phi_0(t) + jb(t - \tau_1)\left(\frac{\pi t}{2T} + \phi_1(t)\right) + j\theta(t - \tau_1)\right]. \end{aligned} \tag{3.13}$$

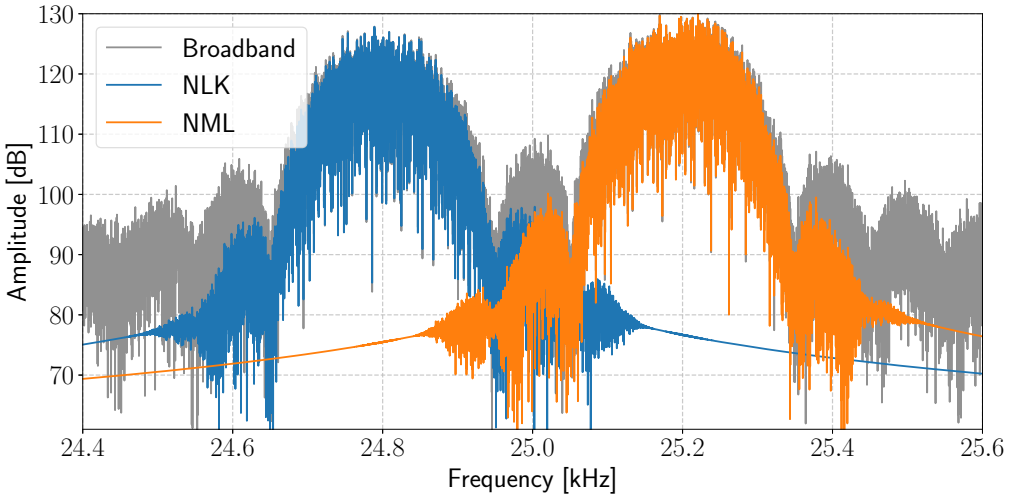

Figure 3.2: Example of NLK's (24.8 kHz) upper sideband and NML's (25.2 kHz) lower sideband interfering with each other.

A factor of 2 is included inside the argument of the lowpass filter to account for half of the signal's energy being placed outside the lowpass filter due to the complex mixing. The amplitude of the MSK is then found by taking the magnitude of the baseband signal

$$A(t) = |z(t)|\,. \tag{3.14}$$

Lowpass filtering is sometimes tricky when two transmitters operate at frequencies close to each other. Figure 3.2 is an example of NLK's (24.8 kHz) upper sideband and NML's (25.2 kHz) lower side band interfering with each other. For MSK demodulation we want to isolate the desired MSK signal from all other signals while retaining as much spectral energy of the MSK signal as possible. Hence, care must be taken when designing a lowpass filter. This work uses a 1001-tap Hanning window filter with a 3-dB cutoff at ~155 Hz, shown in Figure 3.3, which gives a good trade-off between computational burden and band rejection.

As shown before, an MSK signal will alternate between an upper and lower frequency through the use of $b(t)$, and $\theta(t)$ will alternate between 0 and π to maintain phase continuity.

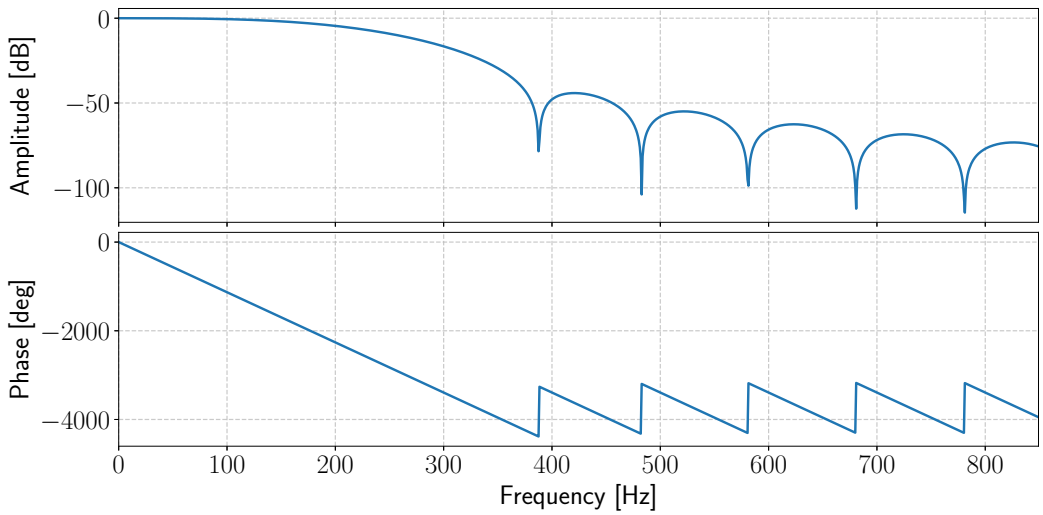

Figure 3.3: Response of lowpass filter used to isolate 100 Hz bandwidth narrowband signals from broadband data.

The $\theta(t)$ term is removed by squaring the baseband signal

$$[z(t)]^2 = [A(t)]^2 \exp\left[j2\phi_0(t) + jb(t-\tau_1)\left(\frac{\pi t}{T} + 2\phi_1(t)\right) + j2\theta(t-\tau_1)\right] \tag{3.15}$$

where $2\theta(t)$ is now 0 or 2π. With the baseband signal squared, $b(t)$ will give either a positive frequency of $1/2T$ or a negative frequency of $-1/2T$ (i.e., $b(t)$ will take on the value ± 1), and the respective phase $2\phi_\pm$ is found by mixing down the $\pm 1/2T$ frequency to baseband and integrating over some period $t_0 \leq t \leq t_1$, where $b(t)$ is constant,

$$2\phi_\pm = \angle\left\{\frac{1}{t_1 - t_0}\int_{t_0}^{t_1} \frac{[z(t)]^2}{[A(t)]^2} e^{\mp j\pi t/T}\, \mathrm{d}t\right\}. \tag{3.16}$$

Note that the normalization of $[z(t)]^2$ by $[A(t)]^2$ helps de-emphasize impulsive energy (such as sferics) that is not filtered out by the lowpass filter but is not part of the MSK signal. In practice (3.16) is calculated over many bits (usually one seconds worth) to acquire an average value for $\phi_\pm$, which also mitigates effects from impulsive noise.

Now that the phase for the upper and lower frequency for the squared baseband signal

has been found, a relationship between $\phi_\pm$, the carrier phase, and the clock phase is formed

$$\begin{aligned}\angle[z_\pm(t)]^2 &= 2\phi_0(t) \pm 2\phi_1(t) \pm \frac{\pi t}{T} \\ &= 2\phi_\pm(t) \pm \frac{\pi t}{T},\end{aligned} \tag{3.17}$$

giving

$$\begin{aligned}\phi_+(t) &= \phi_0(t) + \phi_1(t) + m\pi \\ \phi_-(t) &= \phi_0(t) - \phi_1(t) + n\pi,\end{aligned} \tag{3.18}$$

where $m, n \in \mathbb{Z}$ and have been included to account for the unknown phase wrapping of $\phi_\pm$. It is not uncommon for $\phi_\pm$ to experience phase wrapping when tracking phase changes for long periods. An instantaneous phase jump in $\phi_\pm$ that is close to a multiple of π is most likely caused by this phase wrapping, and m or n can be adjusted to compensate for the phase jump so that phase continuity is maintained. Rearranging (3.18) shows that the carrier phase is proportional to the sum of the upper and lower phase

$$\hat{\phi_0} = \frac{(\phi_+ - m\pi) + (\phi_- - n\pi)}{2}, \tag{3.19}$$

where the $\hat{}$ symbol is used to denote an estimated value. This result is not surprising, considering carrier phase is a proxy for phase velocity, and (3.19) is really just finding the average phase between the upper and lower frequencies. Clock phase is found to be proportional to the difference of the upper and lower phase

$$\hat{\phi_1} = \frac{(\phi_+ - m\pi) - (\phi_- - n\pi)}{2}, \tag{3.20}$$

and using (3.11) the bit transition delay is

$$\hat{\tau}_1 = \frac{-2T\hat{\phi_1}}{\pi}. \tag{3.21}$$

Again this result is expected because clock phase is a proxy for group velocity, and (3.20)

is half the difference of the upper and lower phase.

The final piece in demodulating the MSK signal involves finding the in-phase and quadrature symbol sequence, which is straight forward now that we have the symbol transition delay and know exactly when each symbol occurs. Using (3.13) and (3.7), the baseband signal is written into its in-phase and quadrature form

$$\begin{aligned} z(t) = A(t) \Bigg[& a_I(t - \tau_1(t)) \cos\left(\frac{\pi t}{2T} + \phi_1(t) + \phi_0(t)\right) - \\ & j a_Q(t - \tau_1(t)) \sin\left(\frac{\pi t}{2T} + \phi_1(t) + \phi_0(t)\right) \Bigg]. \end{aligned} \tag{3.22}$$

The in-phase signal $a_I(t)$ has transition boundaries at $t = \tau_1 + T + 2Tk$ for $k \in \mathbb{Z}$. Similarly, the quadrature signal $a_Q(t)$ symbol transitions occur at $t = \tau_1 + 2Tk$. A matched filter method is used to estimate the in-phase and quadrature symbols for each index k,

$$\begin{aligned} \tilde{a}_I[k] &= \frac{1}{T} \int_{\hat{\tau}_1+T+2Tk}^{\hat{\tau}_1+T+2T(k+1)} \Re\left\{ \frac{z(t)}{A(t)} e^{-j\hat{\phi}_0} \right\} \cos\left(\frac{\pi t}{2T} + \hat{\phi}_1\right) \mathrm{d}t \\ \tilde{a}_Q[k] &= \frac{1}{T} \int_{\hat{\tau}_1+2Tk}^{\hat{\tau}_1+2T(k+1)} \Im\left\{ \frac{z(t)}{A(t)} e^{-j\hat{\phi}_0} \right\} \sin\left(\frac{\pi t}{2T} + \hat{\phi}_1\right) \mathrm{d}t \,. \end{aligned} \tag{3.23}$$

The bit stream of an MSK signal b_k is defined as a discrete uniform distribution of ± 1, and each element in the sequence is independent from all other elements. It then follows that the in-phase and quadrature symbol streams are independent and identically distributed discrete random variables. Therefore, the digital in-phase and quadrature symbol streams are found by using the signum function on each discrete analog symbol estimate

$$\begin{aligned} \hat{a}_I(t - \hat{\tau}_1) &= \sum_k \operatorname{sgn}\left(\tilde{a}_I[k]\right) \; p_{2T}(t - \hat{\tau}_1 - T - 2Tk) \\ \hat{a}_Q(t - \hat{\tau}_1) &= \sum_k \operatorname{sgn}\left(\tilde{a}_Q[k]\right) \; p_{2T}(t - \hat{\tau}_1 - 2Tk), \end{aligned} \tag{3.24}$$

and the ambiguous value of $\operatorname{sgn}(0)$ is chosen to be either ± 1.

At this point, all the necessary information about the MSK signal has been extracted

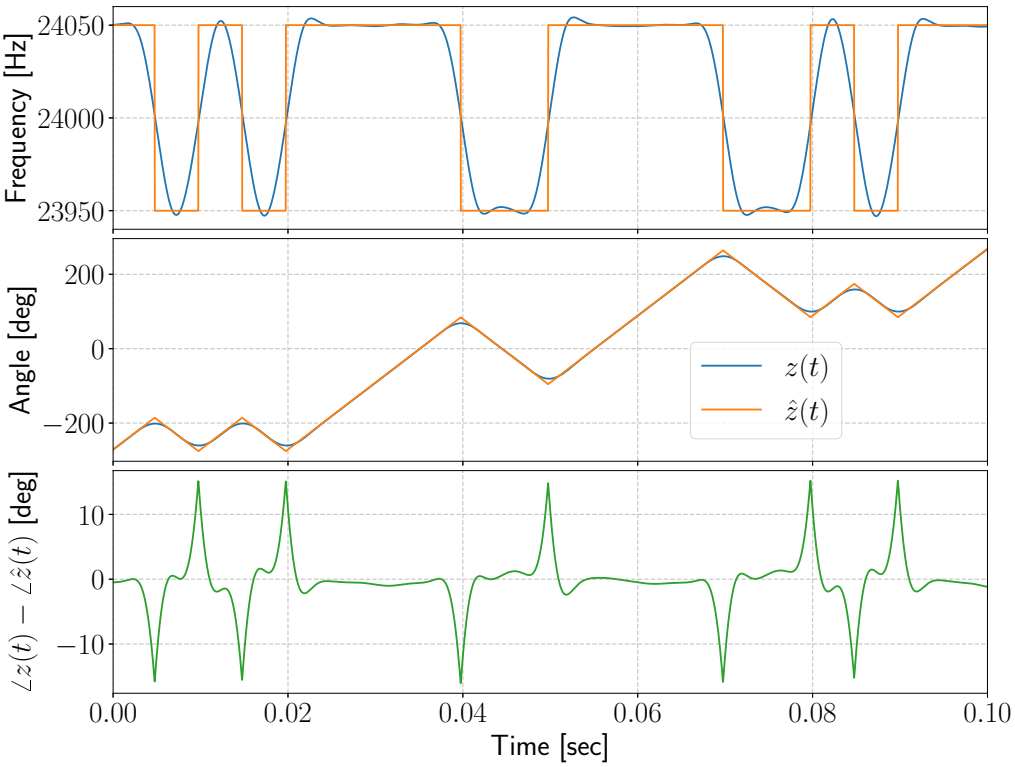

Figure 3.4: Comparison of a reconstructed MSK signal from the NAA (24 kHz) transmitter to the originally recorded MSK signal.

and the phase trellis is reconstructed as

$$\hat{\phi}_{\text{MSK}}(t-\hat{\tau}_1) = \hat{b}(t-\hat{\tau}_1)\frac{\pi t}{2T} + \hat{\phi}_1 + \hat{\theta}(t-\hat{\tau}_1), \tag{3.25}$$

and the estimate for the MSK signal is

$$\hat{z}(t) = A(t)\, e^{j\hat{\phi}_{\text{MSK}}(t-\hat{\tau}_1)+j\hat{\phi}_0}. \tag{3.26}$$

Figure 3.4 is an example MSK signal recording of the NAA (24 kHz) transmitter. The recorded signal $z(t)$ is extracted using (3.13) and the reconstructed signal $\hat{z}(t)$ is found using (3.26). The top panel shows the instantaneous frequency

$$f(t) = \frac{1}{2\pi}\,\frac{\mathrm{d}\left[\angle z(t)\right]}{\mathrm{d}t} \tag{3.27}$$

of the recorded and reconstructed signals, which are alternating between the frequencies 24.05 kHz and 23.95 kHz ($\omega_c \pm \pi/(2T)$). Notice that the reconstructed signal has an instantaneous frequency change between the upper and lower frequencies, while the recorded signal has a noticeable transition period between the two frequencies. There are two important reasons for the slow transition in the observed signal. First, a lowpass filter is applied in (3.13) so the signal is band-limited and cannot be a perfect square signal, using a wider lowpass filter reveals that the recorded MSK signal does have a shorter bit transition period than shown in the figure, but increasing the filter passband also introduces a significant amount of unwanted noise. Second, the transmitter itself has bandwidth limitations which arise from the transmitter's electronics and the fact that the transmitter is electrically small. Non-linear filtering from EIWG propagation does have a small effect [*Koh et al.*, 2018], but the problem here is based on limited bandwidth and not the very small dispersion between two frequencies separated 100 Hz apart.

Performance of the MSK signal reconstruction is visually seen in the bottom two panels of Figure 3.4, where the middle shows the angle of the recorded $\angle z(t)$ and reconstructed $\angle \hat{z}(t)$ signals, and the bottom panel shows the difference between the two angles. It is seen in the middle panel that the angle of the recorded and reconstructed signals overlap each other very well, which shows that $\hat{\phi}_0$ and $\hat{\phi}_1$ were estimated well. Any error in estimating $\hat{\phi}_0$ would add an offset to the angle of the reconstructed signal, but would not affect the instantaneous frequency. An error in estimating $\hat{\phi}_1$ would result in a time shift of the reconstructed signal and would be seen in both the instantaneous frequency and the angle of the signal.

An estimate for the carrier phase and clock phase is found using (3.19) and (3.20), respectively, but these estimates are often taken over a relatively long period (usually one second) which may be too long of an integration period when using carrier phase as a means to remote sense fast temporal changes in the ionosphere. A higher fidelity time series estimate of the carrier phase is found by removing the modulated portion of the observed MSK signal

$$A(t)e^{j\hat{\phi}_0(t)} = z(t)e^{-j\hat{\phi}_{\mathrm{MSK}}(t-\hat{\tau}_1)}. \tag{3.28}$$

Bit transition errors shown in Figure 3.4 will become more apparent as the resolution of $\hat{\phi}_0(t)$ becomes more fine, say 100,000 samples per second. These errors can be mitigated by filtering (3.28). Historically, $\hat{\phi}_0(t)$ has been lowpass filtered and resampled to a sampling period of 0.02 ms for *high resolution* narrowband data and the one second carrier phase estimate from (3.19) has been used for *low resolution* narrowband data. Using a 0.02 ms sampling period gives a good trade-off between mitigating bit transition problems while still giving high enough resolution to capture fast ionospheric disturbances such as Early/fast events.

3.1.3 Coherent MSK Removal

Another reason to estimate the phase trellis of an MSK signal is to coherently remove an MSK signal from broadband data. *Said* [2009] and *McCormick et al.* [2018] demonstrate the utility of coherently removing MSK signals from broadband data for sferic analysis, and show that the coherent removal of MSK signals, presented here, outperforms the removal of MSK signals through notch-filtering. Equation (3.13) is used to isolate the MSK signal $A(t)x(t)$ from broadband data $X(t)$, but we now want to estimate a broadband signal $\hat{X}(t)$ that does not contain the reconstructed MSK signal

$$A(t)\hat{x}(t) = A(t)\cos\left[\omega_c t + \hat{\phi}_{\text{MSK}}(t - \hat{\tau}_1) + \hat{\phi}_0(t)\right] \tag{3.29}$$

but still retains transient sferic information in the same frequency band as the MSK signal

$$\hat{X}(t) = X(t) - A(t)\hat{x}(t). \tag{3.30}$$

Figure 3.5 is an example of coherent MSK removal from broadband data. The top left panel (a) shows the amplitude of the broadband data in the frequency domain. The curve labeled *Original* represents the unaltered broadband data that is recorded at a receiver and shows the prominent peaks that are due to the NLK and NML transmitters. Using the technique from (3.30), the NML MSK signal is removed from the broadband data and plotted with the label *NML Removed*. We can measure the spectral correlation between the

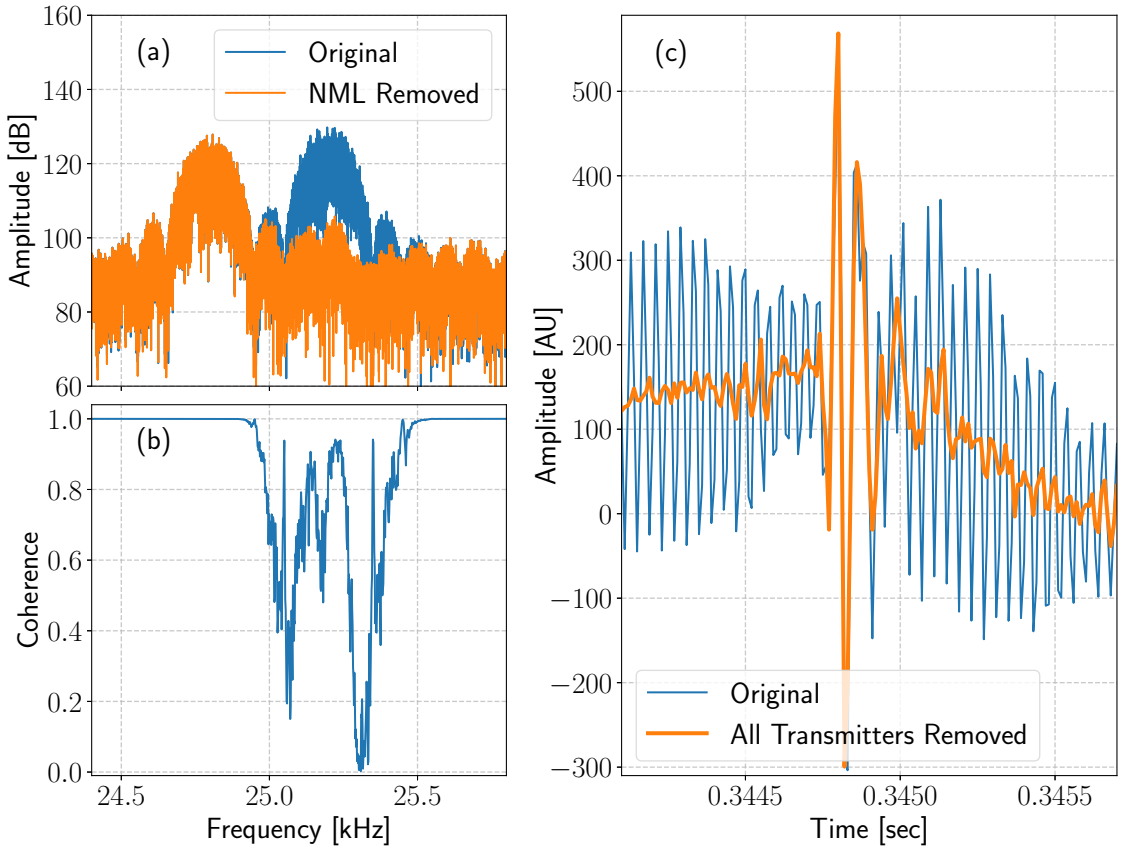

Figure 3.5: Coherent MSK removal example. (a) Broadband data in the frequency domain showing the NLK (24.8 kHz) and NML (25.2 kHz) transmitters together (blue) and the removal of the NML (orange). (b) Coherence of the raw broadband data and broadband data with NML removed. (c) Comparison of a recorded sferic before (blue) and after (orange) the NLK, NML, and NAA (24 kHz) were subtracted from the broadband data.

original and modified broadband data using the coherence function

$$C_{gh}(f) = \frac{|S_{gh}(f)|^2}{S_{gg}(f)S_{hh}(f)}, \tag{3.31}$$

where

$$S_{gh}(f) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(\tau)y(\tau + t)\mathrm{d}\tau \right] e^{-j2\pi ft}\mathrm{d}t \tag{3.32}$$

is the cross-spectral density between the arbitrary time series functions $g(t)$ and $h(t)$, and $S_{gg}(f)$ and $S_{hh}(f)$ are the autospectral densities. The coherence function shows that the two broadband data sets are well correlated outside ±300 Hz of NML's center frequency,

which is good because we want to isolate the removal of NML without interfering with NLK's signal. The right panel (c) shows a relatively weak sferic before (*Original*) and after (*All Transmitters Removed*) the removal of NLK, NML, and NAA. The sferic is visible without the removal of the transmitters, but some of the sferic features, especially the tail end, are hidden by the high energy of the transmitters. Removal of the MSK signals reveals a more complete picture of the fast transient sferic details that were hidden by the high energy MSK signals.

3.2 2-Channel MSK Demodulation

Electromagnetic boundary conditions dictate that the signal strength of VLF waves propagating in the EIWG at Earth's ground are strongest in the vertical direction for E-fields and strongest in the horizontal direction for B-fields, hence most ground-based antennas used in VLF remote sensing are vertical E-field antennas and horizontal B-field antennas. Orientation of a vertical E-field antenna, such as monopole antenna, is simple because it is omnidirectional in the horizontal sense. Orientation of a horizontal B-field loop antenna is a little more complicated, because its beam pattern will have two peaks and two nulls located in the horizontal plane. Therefore, a single loop antenna can be orientated to point towards a particular bearing, but the performance of the antenna will be significantly degraded for received signals that are close to the nulls of the antenna. To get around the null problem, two co-located orthogonal loop antennas are used so that the beam pattern peak of one antenna coincides with the null of the other orthogonal antenna.

A problem arises when most of an MSK signal's energy exists on one antenna channel and the other channel is left with a poor MSK signal to noise ratio (SNR). MSK demodulation on the high SNR channel will perform well, while demodulation on the low SNR channel will perform poorly. From (3.16) it is seen that a low SNR condition will affect the phase estimation of $\phi_{\pm}$ and the accuracy of the carrier and clock phase estimates will be reduced. The problem is further compounded in (3.23), where poor accuracy of the carrier and clock phase will misalign the matched filter integration, causing incorrect symbol estimates (bit-flips) to occur. The following is a method to maximize MSK demodulation performance on

both antenna channels, even when one channel has poor SNR.

The carrier phase ϕ_0 of an MSK signal is estimated by subtracting the phase trellis $\hat{\phi}_{\text{MSK}}(t-\hat{\tau})$ from the phase of the baseband MSK signal $\angle z(t)$, shown in (3.28). We extend the single-channel mathematics, from the previous section, to a two-channel synchronized demodulation method by estimating the symbol streams, $\hat{a}_{I,\text{M}}(t)$ and $\hat{a}_{Q,\text{M}}(t)$, and clock phase $\hat{\phi}_{1,\text{M}}$ on a single channel with maximum SNR (subscript M denotes the digitally rotated channel that maximizes the SNR of the MSK signal), and use these estimates to find $\phi_0(t)$ on the two original channels.

The first step in 2-channel MSK demodulation is to digitally rotate the antennas so that one channel has maximum SNR over the spectral band of interest (it should be that the maximum SNR channel is pointing very close to the bearing of the transmitter). Then the phase trellis $\hat{\phi}_{\text{MSK,M}}$ is estimated using the single-channel method on the maximum SNR channel. From here we want to estimate the carrier phase for two orthogonal channels, both with good SNR. To do this we rotate the baseband signal $z_{\pm 45}(t)$ to be $\pm 45°$ off the direction of maximum SNR, which closely distributes the energy of the MSK signal and noise equally across both channels.

Carrier phase estimates are then found by subtracting the maximum SNR phase trellis from the two baseband MSK signals

$$\hat{\phi}_{0,\pm 45}(t) \approx \angle[z_{\pm 45}(t)] - \hat{\phi}_{\text{MSK,M}}(t-\hat{\tau}_1). \tag{3.33}$$

Similar to (3.16), we also want to find an average carrier phase over some interval $t_0 \leq t \leq t_1$

$$\hat{\phi}_{0,\pm 45} \approx \angle \left\{ \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \frac{z(t)}{|z(t)|} e^{-j\hat{\phi}_{\text{MSK,M}}(t-\hat{\tau})} \mathrm{d}t \right\}, \tag{3.34}$$

so that an estimate of the baseband signal, similar to (3.26), can be made

$$\hat{z}_{\pm 45}(t) = A_{\pm 45}(t)\, e^{j\hat{\phi}_{\text{MSK,M}}(t-\hat{\tau}_1) + j\hat{\phi}_{0,\pm 45}}. \tag{3.35}$$

With the two orthogonal channels synchronized, $\hat{z}_{\pm 45}(t)$ may be rotated to any desired direction without the consequence of low SNR on one channel causing poor demodulation

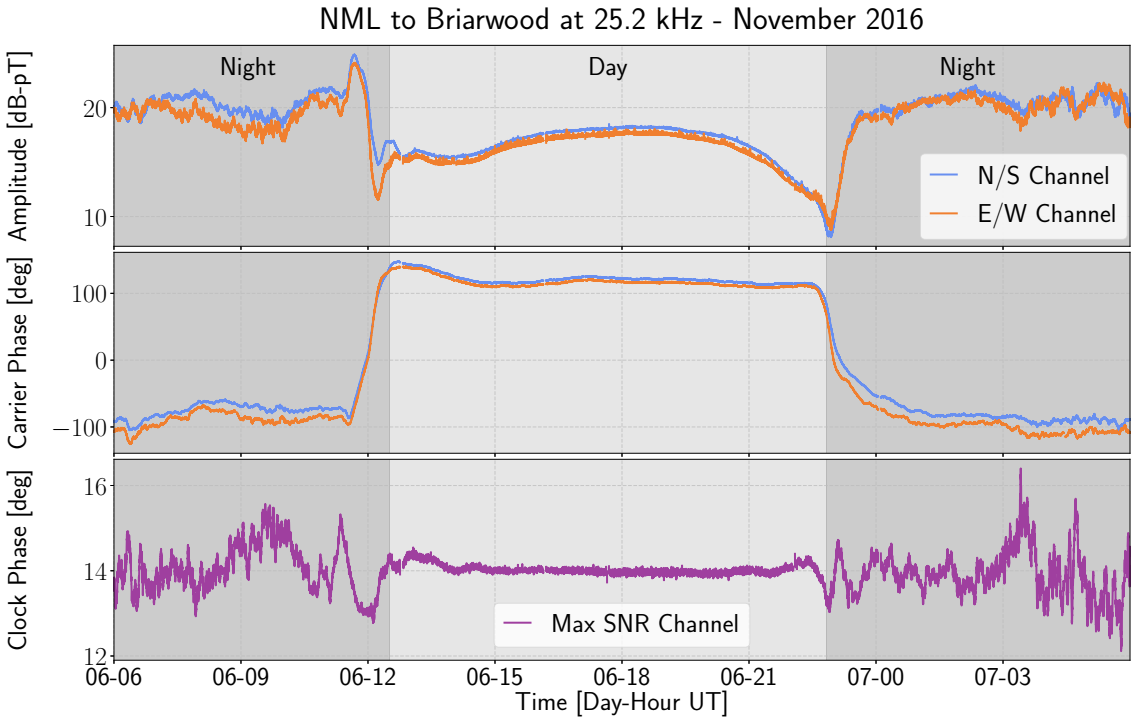

Figure 3.6: Synchronized MSK demodulation of the broadband data measured at Briarwood, tuned to NML's center frequency of 25.2 kHz. The amplitude, carrier phase, and clock phase are shown over a full day for both the N/S and E/W channels. These diurnal curves are typical during quiet ionospheric conditions.

performance.

Figure 3.6 is an example of the 2-channel demodulation algorithm for the NML transmitter being received at Briarwood. In this example, the $\hat{z}_{\pm 45}(t)$ channels have been rotated back to receiver's original N/S and E/W facing channels. The top panel shows the classic and well-known diurnal curves for amplitude data on the N/S and E/W channels. Night-to-day and day-to-night transition periods, known as the day/night terminators, form distinctive nulls and peaks in the diurnal curves and are easily seen. Duration of the observed day/night terminator in the VLF signal is about the time it takes for the sunrise or sunset terminator to traverse the entire transmitter-receiver path. Since the change in longitude along this path is relatively small, the terminator effects are only seen for a short amount of time. The middle panel shows the carrier phase, for both channels, advance and regress from the diurnal variations of the D region ionosphere. During the daytime, the carrier

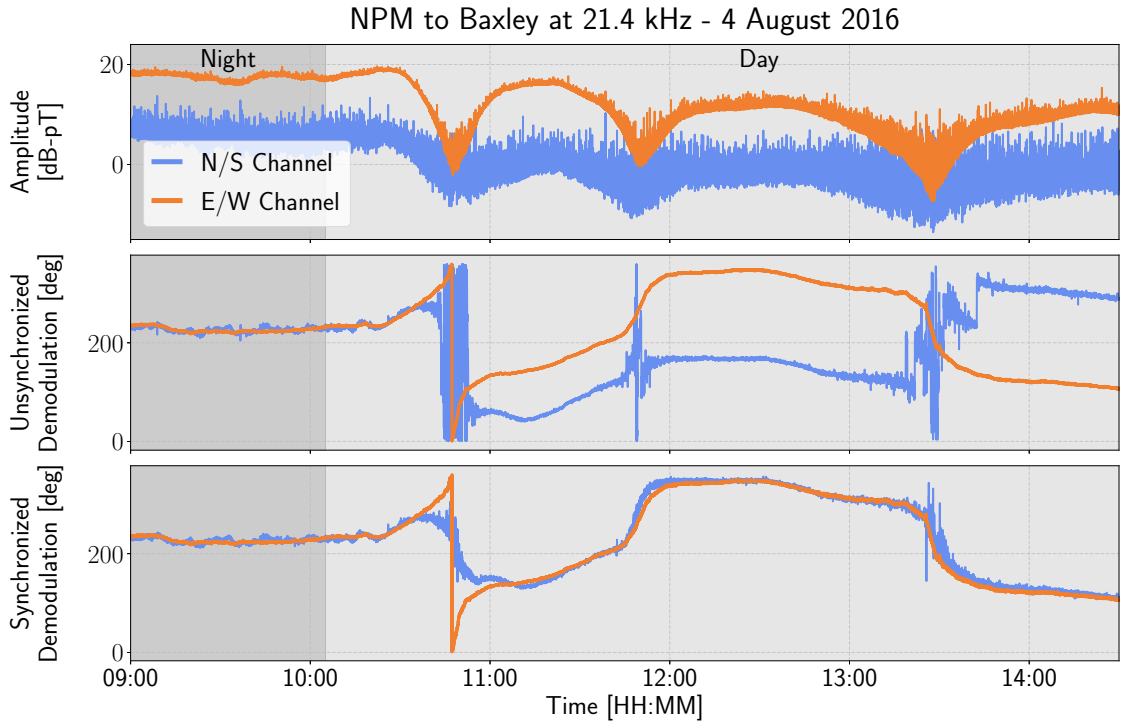

Figure 3.7: Comparison of the unsynchronized and synchronized MSK demodulation algorithm over a long west-to-east path (Hawaii to Georgia). Amplitude is shown in the top panel and has multiple peaks and nulls due to the night-to-day terminator. The unsynchronized carrier phase and synchronized carrier phase is drawn in the bottom two panels, respectively. The nighttime to daytime transition is defined as the point in time where the terminator is directly above the receiver.

phase is smooth, and a small offset is maintained between the two channels. Nighttime carrier phase is less smooth and the offset between the two channels varies more, due to the turbulent nature of the nighttime ionosphere. The bottom panel shows the clock phase from the maximum SNR channel. Like carrier phase, the clock phase is smooth during the daytime and more chaotic during the nighttime.

Figure 3.7, shows the NPM VLF transmitter, in Hawaii, recorded at Baxley, and demonstrates the synchronized demodulation algorithm's utility. Amplitude on both channels is displayed in the top panel. The path from NPM to Baxley has a large longitudinal span, meaning the night-to-day terminator takes several hours to traverse westward along the transmitter-to-receiver path, and this transition is apparent in the received signal between ~10:30−13:45 UT. Three nulls are seen in the amplitude data on both channels and are the

result of modal interference that is commonly observed in long east-to-west or west-to-east paths [*Pappert and Morfitt*, 1975]. During these nulls, the SNR of the MSK signal drops which makes estimation of the MSK parameters more difficult. In the middle panel, the carrier phase has been extracted from each channel individually, using the original single-channel algorithm. Before the first null occurs, the phase difference between channels is typically within ~5° of each other. During the first null, the N/S channel carrier phase begins to sporadically move, showing that the carrier phase is being improperly estimated due to the low SNR condition. Once the first null has passed, a 90° offset between the two channels appears. This offset is the result of the 90° phase ambiguity that arises from (3.19), meaning the N/S channel SNR is too low for the single-channel demodulation algorithm to determine the correct m and n values necessary for proper phase unwrapping. During the second and third nulls, the carrier phase is again uncertain and more 90° phase shifts occur, compounding the 90° phase ambiguity problem. Carrier phase in the bottom panel is calculated using the 2-channel demodulation algorithm with the symbols and clock phase synchronized. During the first two nulls, the carrier phase is estimated well, and during the third null, nearly all points in time are cleanly tracked. Furthermore, erroneous 90° phase offsets are not induced in the N/S channel after each null, showing that the transmitter carrier phase is phase locked between the two channels, and the same symbol estimates and clock phase are being used on both channels. Interestingly, the offset between the N/S and E/W phase is relatively steady except at the first null, where they deviate from each other. This deviation is the result of the terminator and is discussed with more detail later in this chapter.

3.3 Polarization Ellipses

3.3.1 Polarization Definition and Utility

The use of polarization ellipses is a common technique in electromagnetic remote sensing and communications applications, but it is much less common in the field of VLF remote sensing. Many VLF recordings are made with two loop antennas and simple analog or digital filters that isolate the amplitude of a particular VLF transmitter frequency but do not

calculate the phase via MSK demodulation. These recordings allow some direction-finding capability but cannot distinguish between, for example, a linear polarization and a highly elliptical polarization. Furthermore, an accurate measurement of the transverse magnetic flux density amplitude cannot be made unless one of the two co-located loop antennas is deliberately oriented to observe a given transmitter. Often, the co-located loop antennas are oriented in the N/S and E/W directions, so amplitude and phase are required to digitally rotate the antenna towards a given transmitter. Here, we detail the mathematical description of polarization ellipses and apply it to narrowband VLF transmitter signals after demodulation.

At any given time, the horizontal magnetic flux density of a narrowband beacon can be represented by four quantities: amplitude and carrier phase of each of the N/S and E/W antennas (N/S and E/W antennas are used here, but this discussion is easily extended to any pair of orthogonal antennas). The amplitude A and phase ψ of the two antennas will be denoted as A_{NS}, ψ_{NS}, A_{EW}, ψ_{EW}. These four components are written as two separate complex phasors, which define an ellipse centered at the origin of the complex plane. The polarization ellipse technique has previously been applied to broadband signals to locate the arrival direction of chorus signals re-entering the atmosphere from the space environment at high latitudes [*Gołkowski and Inan*, 2008; *Hosseini et al.*, 2017], as well for direction finding of broadband sferics [*Said et al.*, 2010]. *Maxworth et al.* [2015] also applied aspects of the polarization ellipse to coherent signals emitted by ionospheric ELF/VLF modulated HF heating. A portion of the magnetic flux density also exists in the vertical direction and is not always negligible [*Silber et al.*, 2015]. This work only considers the horizontal magnetic flux density, but the proposed polarization ellipse technique may be extended to include the vertical magnetic flux density, which would then give an ellipsoid.

Figure 3.8 shows the geometry and coordinate system involved in polarization measurements. Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ refer to the east and north directions, respectively. The plane of the N/S loop antenna is parallel to the $\hat{\mathbf{y}}$ axis and has a surface vector pointing in the $\hat{\mathbf{x}}$ direction. Likewise, the plane of the E/W antenna is parallel to the $\hat{\mathbf{x}}$ axis and has a surface vector pointing in the $-\hat{\mathbf{y}}$ direction. The polarity of each loop antenna (i.e., which direction the surface vector points) is determined by the direction the wire is wound. When standing

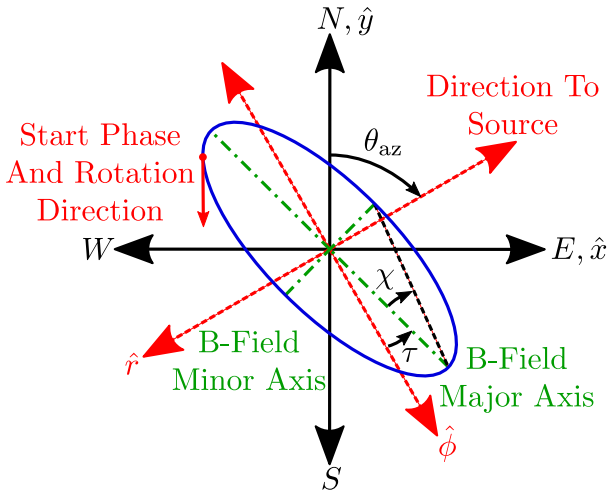

Figure 3.8: Geometric setup for a VLF receiver's antenna orientation (black axes), wave propagation axes (red axes), and polarization ellipse (blue ellipse) of the magnetic flux density. The angle θ_{az} gives the nominal arrival direction from the source. The green markings show the major axis and minor axis of the ellipse, τ is the tilt angle of the ellipse from $\hat{\boldsymbol{\phi}}$, and χ is the ellipticity angle. The red vector on the ellipse shows the start phase and rotation sense of the ellipse.

east of the N/S antenna and facing west, the positive lead of the wire is traced back to the negative lead in a clockwise rotation. The same clockwise wrapping is true when standing south of the E/W antenna and facing north. Amplitude and phase of the received magnetic flux densities are combined to find the total horizontal magnetic flux density

$$\mathbf{B}(t) = A_{\text{NS}} \cos(\omega t + \psi_{\text{NS}})\hat{\mathbf{x}} - A_{\text{EW}} \cos(\omega t + \psi_{\text{EW}})\hat{\mathbf{y}}. \tag{3.36}$$

Note that ψ_{NS} and ψ_{EW} must be correct relative to each other (no 90° ambiguity), so we assume that the 2-channel phase synchronizing demodulation technique, described in the previous section, is used to extract the phase from each channel. Also, the sign of ψ is positive to be consistent with (3.28).

Red axes in Figure 3.8 are shown to align with a hypothetical source located, in this example, northeast from the receiver. The radial direction, $\hat{\mathbf{r}}$, which points radially away from the source, is therefore toward the southwest. Adhering to the definition of a right-hand coordinate system, the azimuthal direction $\hat{\boldsymbol{\phi}}$ is 90° counterclockwise from $\hat{\mathbf{r}}$. The receiver rotation angle θ_{az} is defined as the angle, with clockwise rotation sense, from the

northward direction $\hat{\mathbf{y}}$ to the direction of the transmitter $-\hat{\mathbf{r}}$. A clockwise rotation matrix is used to rotate the magnetic flux density from the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ coordinate system to the $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\phi}}$ coordinate system,

$$\begin{bmatrix} B_r \\ B_\phi \end{bmatrix} = \begin{bmatrix} \cos(\theta_{\text{az}} + \frac{\pi}{2}) & \sin(\theta_{\text{az}} + \frac{\pi}{2}) \\ -\sin(\theta_{\text{az}} + \frac{\pi}{2}) & \cos(\theta_{\text{az}} + \frac{\pi}{2}) \end{bmatrix} \begin{bmatrix} A_{NS} e^{j\psi_{\text{NS}}} \\ -A_{EW} e^{j\psi_{\text{EW}}} \end{bmatrix}, \tag{3.37}$$

where B_r and B_ϕ are the phasor form of the magnetic flux density.

The blue trace in Figure 3.8 shows an example polarization ellipse, it is equivalent to the time domain magnetic flux density $\mathbf{B}(t)$ being drawn on the xy-plane over a full period. There are four properties that together specify the ellipse: the major axis (which is double the semi-major), the tilt angle τ, the ellipticity χ, and the start phase ρ. The polarization ellipse, in general, is not aligned perfectly with the azimuthal direction, but for TM-dominated signals it is often reasonably close. Tilt angle

$$\tau = \frac{1}{2} \tan^{-1} \left[\frac{2\gamma}{1-\gamma^2} \cos(\psi_0) \right], \tag{3.38}$$

where

$$\gamma = \frac{|-B_r|}{|B_\phi|} \tag{3.39}$$

and

$$\psi_0 = \angle \left[\frac{-B_r}{B_\phi} \right], \tag{3.40}$$

is defined as the counterclockwise angle from $\hat{\boldsymbol{\phi}}$ to the major axis direction, with a range of $-90^\circ \leq \tau \leq 90^\circ$. Tilt angle can also be thought of as the error in angle of arrival when using classic magnetic direction finding.

The ellipticity angle

$$\chi = \frac{1}{2} \sin^{-1} \left[\frac{2\gamma}{1+\gamma^2} \sin(\psi_0) \right] \tag{3.41}$$

of an ellipse encapsulates two parameters: how linear (or circular) the ellipse is and the rotational sense of the ellipse. Since the minor axis is by definition smaller than the major axis, the ellipticity angle exists over the interval $-45^\circ \leq \chi \leq 45^\circ$. The rotational sense

of the ellipse is defined by the sign of χ, a positive value ($\chi > 0$) indicates the ellipse is right-hand polarized (or rotating counterclockwise) and a negative value ($\chi < 0$) means the ellipse is left-hand polarized (or rotating clockwise). For the example in Figure 3.8 the *Rotation Direction* vector is showing a right-hand polarization, so χ is positive. The ellipticity angle also describes how linear the ellipse is. When χ is close to zero, the ellipse is highly linear. Conversely, the ellipse is highly circular when the absolute value of the ellipticity ($|\chi|$) is near 45°.

Start phase is the phase difference between the vector parallel to the semi-major axis closest to $\hat{\boldsymbol{\phi}}$, written as $\mathbf{B}(t{=}t_{\mathrm{Maj}})$, and the initial point of the magnetic flux density vector $\mathbf{B}(t{=}0)$, which gives

$$\rho = \omega\, t_{\mathrm{Maj}}. \tag{3.42}$$

Using a counterclockwise rotation matrix, the magnetic flux density phasors along the minor axis and major axis, B_{Min} and B_{Maj}, respectively, can be solved for

$$\begin{bmatrix} B_{\mathrm{Min}} \\ B_{\mathrm{Maj}} \end{bmatrix} = \begin{bmatrix} \cos(\tau) & -\sin(\tau) \\ \sin(\tau) & \cos(\tau) \end{bmatrix} \begin{bmatrix} B_r \\ B_\phi \end{bmatrix}. \tag{3.43}$$

By definition, $B_{\mathrm{Maj}} = |B_{\mathrm{Maj}}|\, e^{-j\omega\, t_{\mathrm{Maj}}}$, which yields $-\angle B_{\mathrm{Maj}} = \omega\, t_{\mathrm{Maj}}$, and allows the start angle to be written as

$$\rho = -\angle B_{\mathrm{Maj}}. \tag{3.44}$$

Start phase is an important metric because it captures transmitter phase changes but is independent of the geometric shape of the ellipse. If the carrier phase changes equally on both the azimuthal and radial channels, then B_r and B_ϕ experience equal phase changes. On the other hand, the shape of the ellipse, whether it is highly linear or circular, is closely connected to the phase difference between the two components. This is useful when dealing with transmitters that have drifting or unstable phase at the source. The phase instabilities are captured in the start phase, but they do not contaminate the geometric properties of the ellipse. Note that since $\mathbf{B}(t)$ forms an ellipse, the start phase is not the same as the geometric angle between $\mathbf{B}(t{=}t_{\mathrm{Maj}})$ and $\mathbf{B}(t{=}0)$, unless $\mathbf{B}(t)$ is circularly polarized.

3.3.2 Angle of Arrival

In Figure 3.8 we define the angle of arrival, for a transmitter signal observed at a receiver, to be θ_{az}. There are two sources of error for the angle of arrival, site errors and receiver errors, both of which can be mitigated through careful post-processing techniques. Site errors are caused by local site anomalies which may include nearby buildings, power lines, cables, and various surrounding terrain [*Mach et al.*, 1986]. Receiver errors come from imperfections in receiver construction and calibration, and can be mitigated during the receiver manufacturing and installation process [*Wood and Inan*, 2004]. Removal of these errors may be accomplished by finding a calibration function f that relates the measured angle of arrival θ_{calc} to the true geometric angle of arrival θ_{truth},

$$\theta_{\text{truth}} = f(\theta_{\text{calc}}). \tag{3.45}$$

Site errors are due to re-radiated signals from conductive materials within the vicinity of the receiver. An incoming VLF wave will produce currents on the conductive material, which will then create secondary waves (re-radiate) that are received by the receiver but at some angle that is different than the incoming VLF wave. Angle of arrival errors from the secondary wave have been shown to be quadrantal errors, which are two-cycle sinusoidal curves with even and odd harmonics [*Miyake et al.*, 1995]. *Hiscox et al.* [1984] shows the effects from the secondary wave scattering can often be approximated to a simple two-parameter function

$$f_{\text{quadrantal}}(A, \phi, \theta_{\text{calc}}) = \theta_{\text{calc}} - A\sin(2\theta + \phi), \tag{3.46}$$

which is just the fundamental frequency of the quadrantal error, but also noted that a more complex equation is needed when quadrantal harmonics are not negligible (see *Passi and López* [1989]). In practice, multiple re-radiating sources exist, and a receiver's site error is the superposition of all quadrantal errors. The sum of these quadrantal errors still produces an error function with a two-cycle fundamental frequency, as shown in (3.46), because the sum of two sinusoids with equal frequencies gives a sinusoid at the same frequency.

Receiver errors, with regard to angle of arrival, are symptoms of imperfect tolerances in the electronics of the receiver and imperfect construction and deployment of the antenna. *Wood and Inan* [2004] developed a three-parameter model motivated by antenna imperfections to correct for receiver errors,

$$f_{\text{antenna}}(\alpha, \xi, \rho, \theta_{\text{calc}}) = \arctan\left[\alpha \frac{\tan(\theta_{\text{calc}})}{\cos(\xi)} - \tan(\xi)\right] + \rho, \tag{3.47}$$

which accounts for the deviation of the N/S antenna from due north ρ, the skew of the two antennas relative to orthogonal ξ, and the ratio of the two antenna surface areas α. Gain differences between the two channels in the electronics of the receiver are removed using an internal calibration method [*Cohen et al.*, 2018a], but any residual gain errors between the electronics of the two channels will also be captured in the α term.

Measured angle of arrival errors are the combination of site errors and receiver errors. Decoupling the two types of errors is challenging, and the choice of using either the quadrantal correction function (3.46) or the antenna correction function (3.47), or both, is a matter of preference and understanding of the driving mechanism for angle of arrival errors at a given site. We continue with our discussion using the *Wood and Inan* [2004] method only, but note that this discussion may be adapted to use the quadrantal method, or both.

Estimation of the three parameters in (3.47) is done by minimizing a cost function related to (3.45),

$$\arg\min_{\alpha,\xi,\rho} \sum_{i=1}^{N} \left[\theta_{i,\text{truth}} - f_{\text{antenna}}(\alpha, \xi, \rho, \theta_{i,\text{calc}})\right]^2, \tag{3.48}$$

over a set of N angle of arrival measurements. Ideally, the set of angle of arrival measurements would be uniformly spaced from 0° to 360°, but too few VLF transmitters exist to give a good spatial distribution for angle of arrival measurements. Instead, lightning generated radio sferics are used to generate the spatial distribution required to solve (3.48).

A few considerations must be made when using sferics to correct for angle of arrival errors. First, the ground truth angles θ_{truth} are found by assuming the radio wave source is the location of a lightning stroke, which is reported by some lightning detection network. Any error in the detection network will ultimately result in θ_{truth} errors. We use the Vaisala

Inc. GLD360 network, which has a median cloud-to-ground stroke location accuracy of ~2.5 km [*Said et al.*, 2010, 2013]. Second, the measured angle of arrival is found using the polarization ellipse tilt angle technique in (3.38), which works best when $B_\phi \gg B_r$ (the dominant TM mode is much larger than all TE modes combined). Tilt angle measurements are most reliable for angle of arrival calibration when sferics have propagated over long daytime paths [*Horner*, 1954], so we filter our selected sferics to be those that originate from 1,000 to 3,000 km away from the receiver and those that have propagated along a total daytime path. Third, anisotropy in the EIWG causes frequency dependent tilt angle changes [*Yamashita and Sao*, 1974a,b]. We calculate the measured angle of arrival by averaging the measured polarization tilt angle $\tau_i(\omega)$ over a small frequency range ($\omega_0 \leq \omega \leq \omega_1$)

$$\theta_{i,\text{calc}} = \theta_{i,\text{truth}} - \frac{1}{\omega_1 - \omega_0} \int_{\omega_0}^{\omega_1} \tau_i(\omega)\, \mathrm{d}\omega \tag{3.49}$$

to minimize the tilt angle frequency dependence while also using a large enough frequency band to reduce tilt angle jitter induced from noise in the measurements.

Angle of arrival error correction for the Briarwood receiver is shown in Figure 3.9 with frequency integration over the intervals 8 to 12 kHz and 20 to 23 kHz. Cloud-to-ground sferics were filtered to be within the days 11 July 2017 to 8 August 2017, between the times of 16:00 to 20:00 UT, and within the distance range of 1,000 to 3,000 km from the receiver, which resulted in over one million sferic waveforms. An advantage for filtering over a large number of days with a small time window each day is that any ionospheric anomalies that persist for only a day or two, say from high solar activity, will be statistically de-emphasized since a large number of days are used. Two different frequency ranges were used to contrast the trade-offs between the use of a lower and higher frequency integration window. Sferic energy typically peaks at frequencies between 5 to 10 kHz [*Le Vine*, 1987], so a frequency range of 8 to 12 kHz is chosen. A range of 20 to 23 kHz is also considered because we are interested in correcting measured VLF transmitter polarization ellipse tilt angles, which are close to 25 kHz. It is important to consider the possibility of site errors and receiver errors having a frequency dependence, which would imply the frequency integration range should be centered around the transmitters' frequencies, but if the range includes the transmitters'

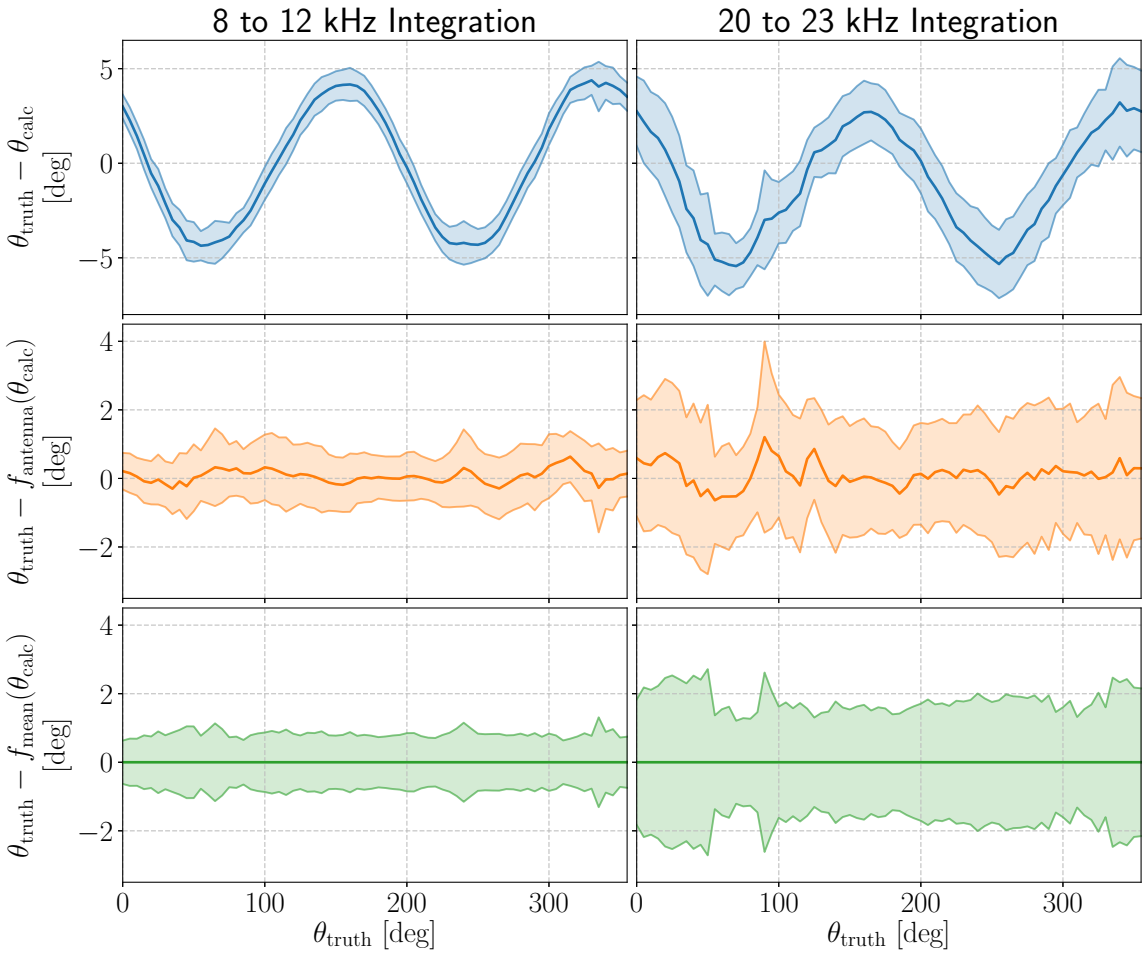

Figure 3.9: Angle of arrival error correction for the Briarwood receiver with frequency integration over 8 to 12 kHz and 20 to 23 kHz. The shaded region represents one standard deviation from the mean error. Measured angle of arrival error with no correction is shown in the top row. The correction method from (3.47) is shown in the middle row. The bottom row is the mean error removal method.

frequencies, then the transmitters themselves may induce errors in the angle of arrival error corrections. Therefore, a frequency range that has no noticeable transmitter energy, but is still close to the VLF transmitters' frequencies, is chosen: 20 to 23 kHz.

The top row in Figure 3.9 shows the measured angle of arrival error with no correction, and the shaded region (in all the panels) represents one standard deviation from the mean error. A two-cycle sinusoid is seen for both frequency ranges, and whether this sinusoid is due to site errors or receiver errors is unknown, but it is obvious that the two-parameter equation (3.46) resembles the angle of arrival error seen here.

There are a few noticeable differences between the no correction error curves for the lower and upper frequency integration ranges. The error curve for the lower frequency range has a higher offset than the upper frequency range, and the two-cycle sinusoid phase of the upper frequency range appears to slightly lag the lower frequency range. These two differences are likely caused by frequency dependence in site errors and receiver errors. Also notice that the standard deviation in the lower frequency range is generally smaller than the upper frequency range, this difference is expected since the lower frequency range has a better sferic SNR.

Error correction using the three-parameter antenna correction method from (3.47) is shown in the second row of Figure 3.9. Best fit parameters are solved for using (3.48), giving lower frequency range parameters of $(\alpha, \xi, \rho) = (0.90, -6.0^\circ, -3.3^\circ)$ and upper frequency range parameters of $(\alpha, \xi, \rho) = (0.94, -6.8^\circ, -4.6^\circ)$. The results show the three-parameter method performs well in correcting the site and receiver errors. Performance for the lower frequency range appears to be better and is most likely due to the smaller standard deviation for the lower frequency range. Although, there is the possibility that the higher frequency range contains stronger high order quadrantal errors, which are not well accounted for in (3.47).

A final method using $f_{\text{mean}}(\theta_{\text{calc}})$, which was not discussed previously, is shown in the bottom row. This method is the subtraction of the uncalibrated mean angle of arrival error, shown in the first row, from θ_{calc}. Hence, the average error for $\theta_{\text{truth}} - f_{\text{mean}}(\theta_{\text{calc}})$ is zero for all angles, but still has the same standard deviation as before. The use of $f_{\text{antenna}}(\theta_{\text{calc}})$ and $f_{\text{mean}}(\theta_{\text{calc}})$ show similar results, and it is difficult to say which method is the best to use. On one hand, $f_{\text{antenna}}(\theta_{\text{calc}})$ is a parametric equation that is derived using antenna geometry and has a physical meaning. On the other hand, $f_{\text{mean}}(\theta_{\text{calc}})$ may capture some physical anomalies, like higher order quadrantal errors, that $f_{\text{antenna}}(\theta_{\text{calc}})$ is not designed to account for. For the remainder of this work, the $f_{\text{mean}}(\theta_{\text{calc}})$ method will be used, but note that switching to $f_{\text{antenna}}(\theta_{\text{calc}})$ would produce almost identical results.

3.3.3 Ellipse Analysis Example

Now that the mathematics for polarization ellipse analysis have been established, we want to see if any interesting ellipse features present themselves that are not easily seen when looking at raw amplitude and phase data. Figure 3.10 is the same 24-hour VLF transmitter dataset from Figure 3.6 reformulated into the polarization ellipse parameters. The top panel shows the semimajor and semiminor lengths. We see that the semimajor is much larger than the semiminor, meaning the ellipse is close to linearly polarized, and is because the TM mode of VLF wave propagation dominates at longer distances. Also, the major axis is relatively steady with time, but the minor axis fluctuates greatly, and at times disappears almost entirely as the signal becomes linearly polarized.

An ellipse can also be decomposed into two counter-rotating circles

$$\begin{aligned} B_{\mathrm{RHCP}} &= \frac{B_\phi - jB_r}{2} \\ B_{\mathrm{LHCP}} &= \frac{B_\phi + jB_r}{2}, \end{aligned} \tag{3.50}$$

commonly referred to as right-hand circular polarization (RHCP) and left-hand circular polarization (LHCP), and these are shown in the second panel. The LHCP signal is slightly stronger for the entire day, except for a short amount of time during the night-to-day terminator. The change in rotation direction can also be seen in the ellipticity data. The ellipticity angle is greater than zero for a short time during the night-to-day terminator, indicating that the rotation direction briefly changed to a counterclockwise rotation. Also, the ellipticity angle shows the ellipse is quite linear during the day but slightly more circular during the night. This result is not surprising since nighttime conditions likely support a higher number of propagating modes.

The bottom panel of Figure 3.10 shows the last two parameters, the tilt angle and start phase. The tilt angle has been calibrated using the $f_{\mathrm{mean}}(\theta_{\mathrm{calc}})$ angle of arrival calibration method and is roughly $+1.5^\circ$ during the daytime, which is reasonable for angle of arrival measurements. The final parameter, start phase, is related to the phase velocity of the signal (along with the transmitter source phase variations), and therefore provides some

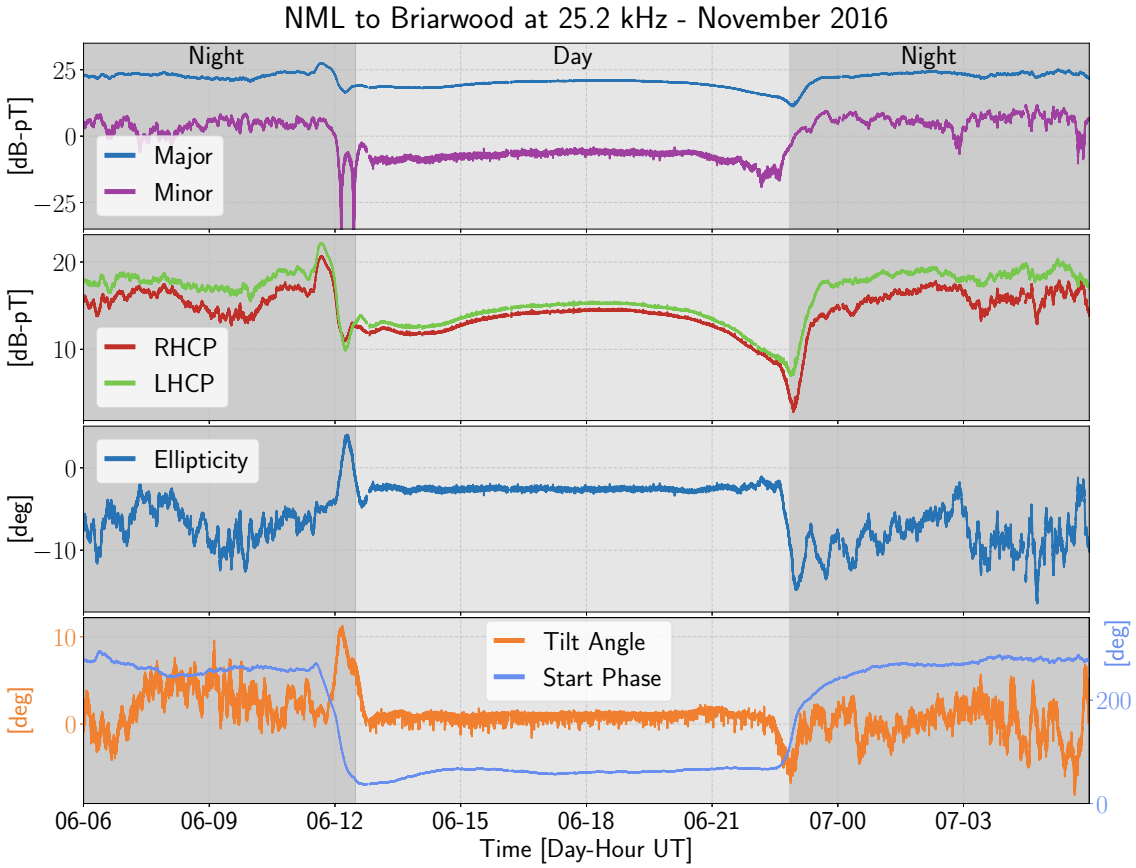

Figure 3.10: Data from Figure 3.6 transformed into ellipse parameters. The first panel shows the semimajor axis and semiminor axis lengths of the time varying ellipse, while the second panel shows the decomposed LHCP and RHCP ellipse components. Ellipticity, shown in panel three, is the rotation sense and how linear (or circular) the ellipse is. The fourth panel shows the ellipse tilt angle and the ellipse start phase.

information on both. If we assume the magnetic flux density is dominated by a single mode and the phase of the transmitter is stable, then the start phase in Figure 3.10 implies the phase velocity is ~0.003c m/s faster during the day. However, these assumptions are not generally true. The quick change in tilt angle and start phase during the terminator periods are most likely the result of mode conversion during the day/night terminators [*Kaiser*, 1968].

Every parameter in Figure 3.10 shows steady and smooth trends during the daytime. Conversely, the night time ellipse parameters are variable and reflect the more chaotic nature of the night time D region ionosphere. As expected, the ellipse parameters are

highly variable during the day/night terminator, showing that the sharp transition region between the daytime ionosphere and nighttime ionosphere greatly affects the propagation of VLF waves traversing the terminator and can significantly increase or decrease the power received from a transmitter.

3.3.4 Phase Stability of VLF Transmitters

Even though VLF transmitters have been used in decades of remote sensing studies, Figure 3.11 shows clear evidence that the phase of at least some VLF transmitters is not always stable. The figure shows the carrier phase and the polarization parameters for the NAA transmitter recorded at Baxley, Georgia. Note that the time period is during the day and no significant solar activity is present. The left column presents the carrier phase measured on the N/S (top) and E/W (bottom) channels, both of which exhibit sharp phase transitions. Though not shown, these phase changes are also seen at other receivers at the exact same time, but only on the NAA transmitter. These sharp phase transitions are certainly a result of an unstable carrier phase at the transmitter and not from transient ionospheric effects. Although these phase transitions are sharp and plainly obvious when looking at phase measurements from multiple sites, it may be easy to mistake one of these transitions as an actual ionospheric disturbance, particularly if only one such sharp transition occurs and if data from only one receiver is considered. It is worth noting that at least some ionospheric disturbances have been known to affect only the phase, and not the amplitude [*Cotts and Inan*, 2007]. This example underscores the importance of considering transmitter source phase transitions in identifying ionospheric events. More importantly, these sharp transitions are likely accompanied by much more gradual drifts in the transmitter source phase that cannot be inferred at the receiver, because it is difficult to decouple phase changes due to slow variations of the ionosphere and phase changes due to slow transmitter phase drifts.

As a consequence, carrier phase measurements are not invariant to transmitter source phase wanderings, whereas amplitude measurements can be more carefully calibrated [*Cohen et al.*, 2010a]. Instead, carrier phase measurements are useful in a relative sense and on short timescales. That is, if we assume the transmitter source phase variations are relatively

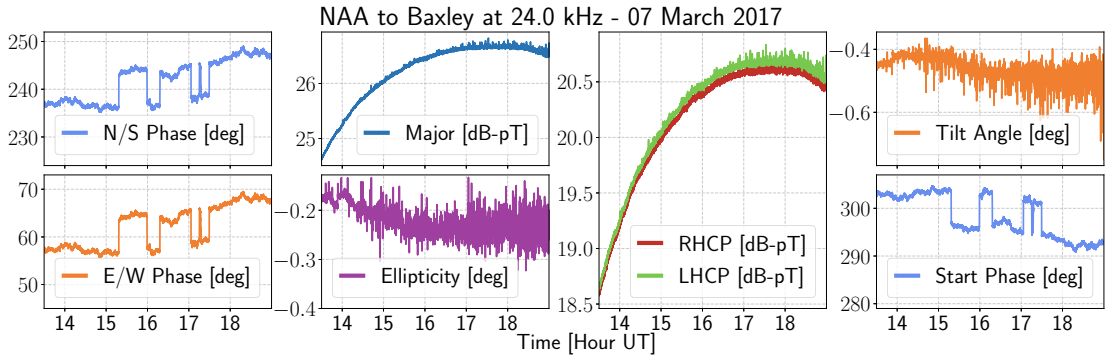

Figure 3.11: Evidence of unstable carrier phase in a VLF transmitter. The carrier phase on both the N/S and E/W channels (left column) has strong step-like changes that are caused by phase instabilities at the transmitter. The ellipse analysis (three right columns) isolates the transmitter phase instability to the start phase, and the other ellipse parameters are uncontaminated by the instability.

slow compared to the time scale of ionospheric changes under investigation, then changes in carrier phase are a reliable measurement to use. However, if studying hours-long trends or day-to-day variations in ionospheric conditions, measured phase changes may be unreliable. *Thomson et al.* [2014] uses a direct approach to the mitigate source phase problem by placing a receiver near the transmitter and measuring the phase, they then use the observed phase from the short path to calibrate the received transmitter phase on the longer paths. Unfortunately, this requires significant effort to maintain sites next to every VLF transmitter. Instead, the polarization ellipse technique can be applied to the narrowband data and the ellipses from each site can be directly compared, even though each site still has an ambiguous phase offset.

The right three columns of Figure 3.11 contain the polarization ellipse parameters. As seen, only the start phase measurement contains sharp phase transitions that are correlated with the carrier phase measurements on the N/S and E/W channels. This shows that the polarization ellipse technique is a useful way to characterize magnetic flux density measurements on two channels even when the source phase is unstable.

3.4 Ionospheric Remote Sensing

We now apply the polarization ellipse technique to narrowband data associated with D region ionospheric disturbances in order to demonstrate features of the received VLF signal that may not be apparent in the single-channel amplitude and phase data. The first example is modal changes of VLF propagation due to the day/night terminator, a somewhat predictable and repeatable phenomenon. Three other observation examples are shown with different types of ionospheric perturbations and different ionospheric modification mechanisms: Early/fast, LEP, and solar flare. Each type of perturbation produces a distinctly different set of polarization parameters. Ellipse analysis was also recently used by *Cohen et al.* [2018b] to observe ionospheric perturbations from the 2017 Great American Solar Eclipse. Note that these results should not be generalized since only an example case study of each type of disturbance is presented. Since these phenomena are all dependent on, for example, path geometry of the transmitter-disturbance-receiver, an in-depth statistical study is required to understand how the changes in polarization ellipses link to causative physical mechanisms. At the same time, the markedly different responses of the polarization ellipses for these examples are representative of the value of utilizing the polarization ellipse method.

3.4.1 Modal Changes Due to Terminator Effects

Understanding the day/night terminator and its effect on VLF wave propagation is, in general, a difficult problem. The strong inhomogeneity of the D region along with the sharp change in electron density has made modeling propagation across the terminator challenging. Figure 3.7, which is the signal from the NPM transmitter in Hawaii recorded at Baxley, Georgia, is an example of the complex changes in VLF wave propagation as the terminator traverses the ionosphere. In particular, the first null in Figure 3.7, which occurs at about 10:48 UT and is referred to as *the first null* for the rest of this section, shows a peculiar deviation between the synchronized N/S and E/W phases. The first null occurs about 43 minutes after the terminator passes over the receiver and is well over the great-circle path. It is difficult to tell what this deviation means when analyzing the amplitude

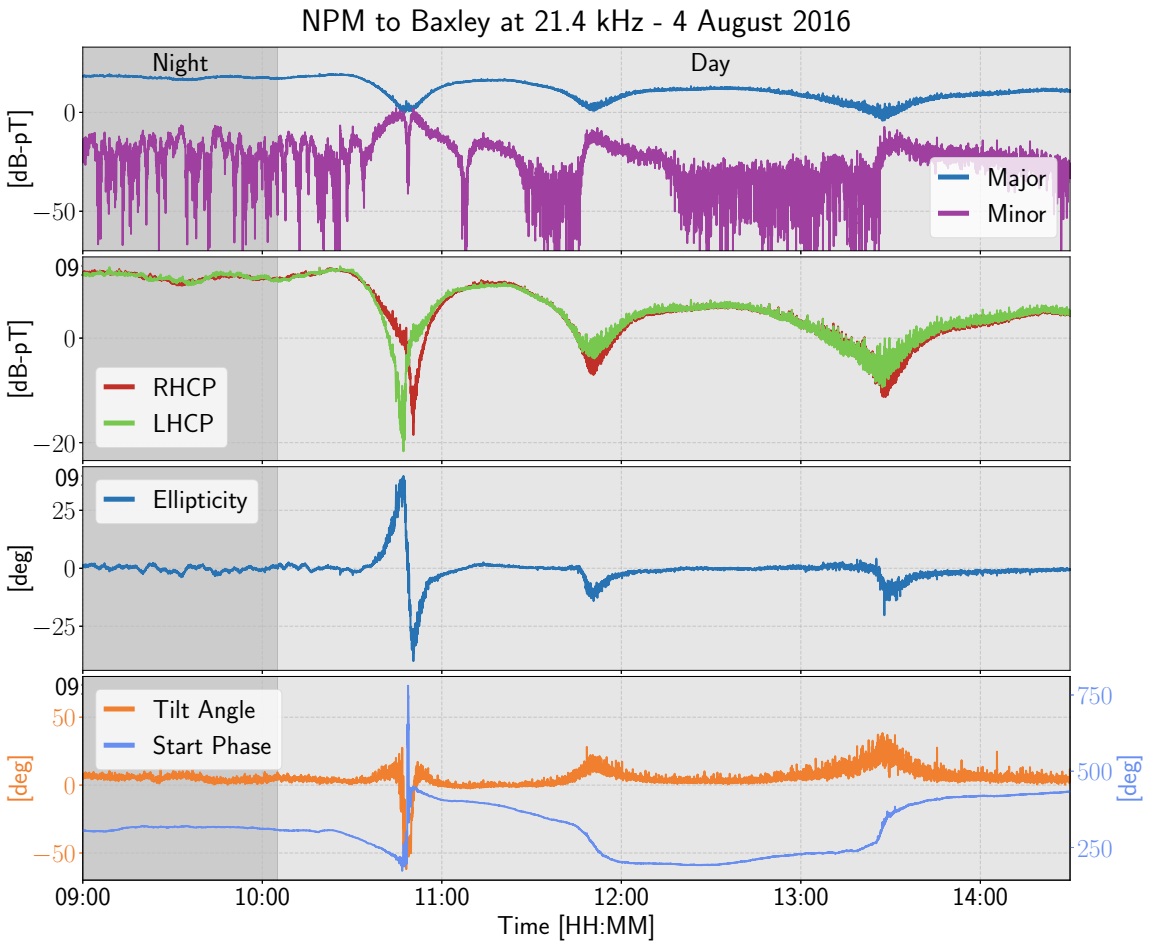

Figure 3.12: Data from Figure 3.7 transformed into ellipse parameters. The phase synchronized E/W channel in Figure 3.7 shows an interesting artifact during the first null, which is also seen in the semimajor curve, shown in panel one. Using ellipse analysis, the artifact is determined to be caused by a separation between the LHCP and RHCP nulls (about 2 minutes before and 2 minutes after the first null, respectively), shown in panel two. A strong change in ellipticity (panel three) and tilt angle (panel four) is also seen during the first null, but the severity in change is not seen during the other two semimajor nulls. Changes in start phase (panel four) are strongest around each semimajor null.

and phase. However, when looking at the first null in terms of ellipse parameters, shown in Figure 3.12, it is obvious that the first null is undergoing a change in elliptical polarization.

The LHCP and RHCP components, in the second panel of Figure 3.12, have nulls at different times. The LHCP component of the ellipse reaches its minimum about 2 minutes before the major axis null, but the RHCP portion reaches a null about 2 minutes after the major axis null. Hence, the RHCP component dominates over the LHCP component for a

brief period just before the major axis null, whereas the LHCP component dominates for a brief period after the major axis null. The signal then returns to being almost linearly polarized. This transition can also be seen in the ellipticity curve. Before the first null, the ellipticity is much greater than zero, meaning that it is highly circular and right-hand polarized. Then, as the first null passes, the ellipticity quickly becomes much less than zero, which indicates the ellipse is again highly circular, but this time left-hand polarized. Note that the quick change, during the first null, in ellipticity angle and start phase is not instantaneous and is not caused by phase wrapping, instead, the change from one extreme to the other is quick, smooth, and continuous.

The cause for the VLF wave to transition from LHCP to RHCP during the first null is not fully understood. Mode conversion from the night to day transition does affect the RHCP and LHCP components differently, due to ionospheric anisotropy from Earth's background magnetic field, which may also explain why the LHCP component is significantly stronger in the second and third nulls. Also, given that the terminator forms a sharp gradient in the D region electron density profile, the receiver may be measuring signals that are obliquely forward scattered from the terminator (i.e., reflecting off the terminator at a region which is not on the great-circle path, causing multipath propagation). It's worth noting that this flip in the rotational sense only occurs for the first null, when the day/night terminator has drifted a few hundred km past the receiver toward the transmitter. We will show in the next subsection an Early/fast event, which is generally caused by ionospheric scattering, that also produces an increase in circular polarization and a strong change in tilt angle. The changes for the case study Early/fast event are similar, but smaller in scale to what is seen here during the first null.

3.4.2 Early/Fast Event

Early events are VLF perturbations caused by powerful or intense lightning near the transmitter-receiver path. Early events result from the coupling of a lightning stroke to the bottom D region ionosphere which in turn creates localized heating and ionization of the lower ionosphere [*Inan et al.*, 1996]. Proposed physical mechanisms include scattering from a sprite halo [*Moore et al.*, 2003], scattering off a sprite body [*Dowden et al.*, 1997],

ionization from the electromagnetic pulse or *elve* [*Mika et al.*, 2006; *Haldoupis et al.*, 2013], quasi-electrostatic quiescent heating [*Pasko et al.*, 1996; *Kabirzadeh et al.*, 2017], and heating from intracloud pulses [*Marshall et al.*, 2008]. Recovery for Early events is typically 10−100 seconds but sometimes much longer [*Cotts and Inan*, 2007], and may occur as far away as 300 km from the path [*Salut et al.*, 2013]. Early events are seen in narrowband data because the perturbed region of the ionosphere scatters a signal that is not on the great-circle path towards the receiver [*Dowden et al.*, 1996]. The receiver measures the superposition of the ambient signal that propagates directly along the great-circle path and a scattered signal from the perturbed region. The recorded amplitude and phase either increases or decreases depending on how the two signals interfere with each other [*Moore et al.*, 2003].

Figure 3.13 shows a typical Early/fast event, with a rapid ($<$ 20 ms) onset duration. The National Lightning Detection Network (NLDN) [*Cummins and Murphy*, 2009] measured the peak current of the stroke to be +156 kA and located the stroke at 43.05°N and 86.03°W. Each column shows the measured narrowband data at Delaware from the NLK and NML transmitters, respectively. This particular event occurred along two transmitter-to-receiver paths which almost exactly overlap.

The top two rows of Figure 3.13 show the amplitude and phase on each channel. The data are recorded at both 1-second and 20 ms time resolutions, but our plots here show only the low-resolution since they are less noisy. We observe the typical response for an Early/fast event: a quick ($<$ 20ms) onset duration followed by a slower recovery. The third row shows the start phase at both receivers. The curve of the start phase closely follows the phase on the N/S and E/W channels, but with opposite polarity. This strong change in start phase is the result of carrier phase changes with respect to the sum of all modes from the direct path and the scattered path.

Narrowband data is inherently susceptible to impulsive noise from, in particular, lightning generated sferics. This impulsive noise is reduced by applying a 3-point median filter to the narrowband data before using the ellipse analysis. The polarization ellipse results are shown in the fourth row. The green ellipse represents the narrowband signal before the causative stroke, and the red ellipse is after the onset of the Early/fast event. The panels in

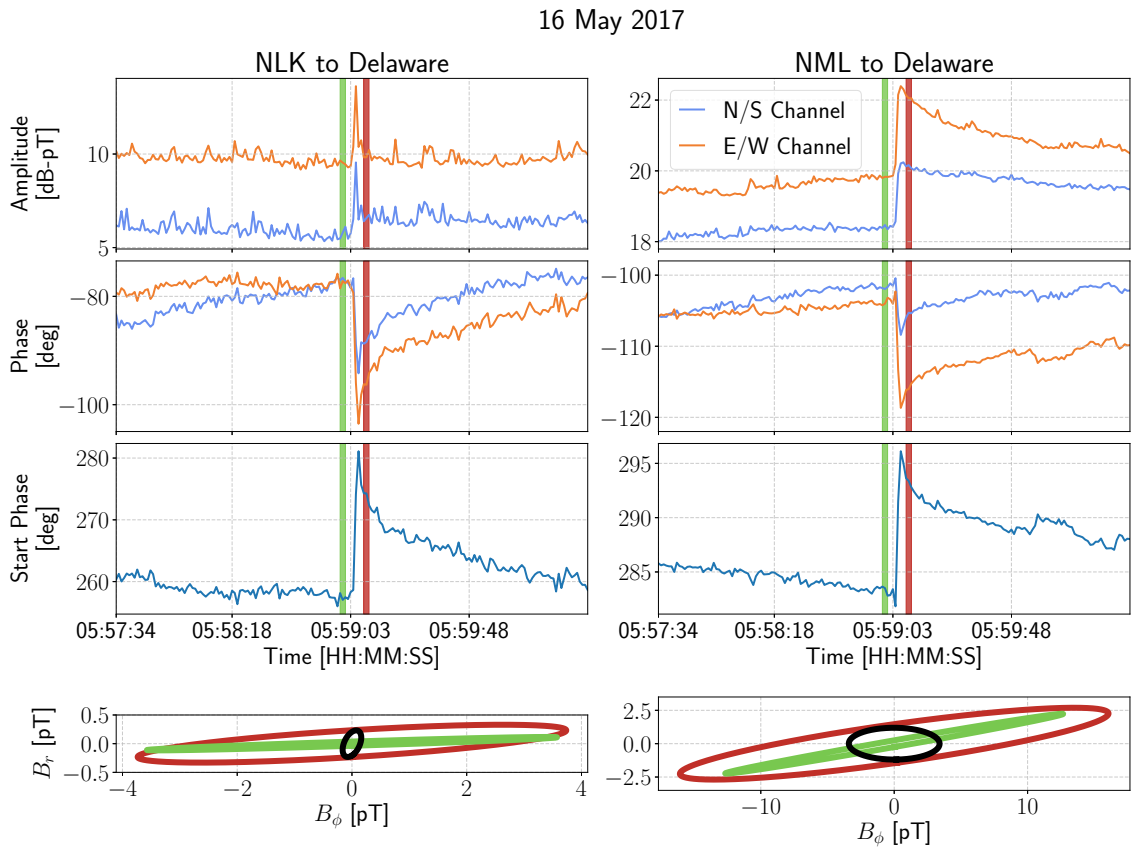

Figure 3.13: Polarization ellipse technique used on an Early/fast event, recorded at a single receiver from two different transmitters. Amplitude, carrier phase, and start phase is shown in the first three rows, respectively. The bottom panels show the ellipses before (green) and after (red) the event occurs, along with the scatter field ellipses (black). Vertical green and red bars in the top three rows show the time periods that the pre-onset ellipses and post-onset ellipses are windowed over.

the top three rows show red and green vertical bars to indicate the time period over which the ellipse parameters were determined. By subtracting the ambient signal (green) from the disturbed signal (red), we can obtain a scattered field polarization ellipse which is shown in black on the bottom two panels. The scattered field ellipse captures the change in major axis, minor axis, and tilt angle, as a result of the ionospheric disturbance.

For the NLK to Delaware path, the Early/fast event causes a rotation of the ellipse and a significant increase of the semi-major axis. Likewise, the NML to Delaware path also has a change in tilt angle but opposite in direction, and an increase on both the semi-major

and semi-minor components. These changes can also be seen in the scattered ellipses. The increase in the semi-major component means that an increase of VLF power is received at the antenna close to the semi-major direction. This power increase may be due to the scattering effect (i.e., multipath) that Early/fast perturbations produce, since this particular event is relatively close (25.7 km) to the great-circle path.

The observed rotation and ellipticity change of the polarization ellipse are worth highlighting. This effect implies that the Early/fast event has different characteristics if observed independently on each of the two channels, both in terms of the amplitude and phase changes, as well as the recovery event. Past works have observed, for instance, that some Early/fast events are only observed on one channel, or that the recovery time differs on the two antennas. Polarization observations indicate at least observationally why this occurs, and observation of a rotating ellipse may be useful in identifying properties of the scattering region, as well as new scattered modes that are created at the disturbed ionosphere.

3.4.3 LEP Event

Lightning-induced electron precipitation (LEP) events are also driven by lightning strokes, but the ionospheric disturbance is triggered in a different way: lightning produces VLF energy that leaks into the magnetosphere and interacts with geomagnetically trapped particles, causing some electrons to precipitate into the lower ionosphere, ultimately increasing the electron density of the D region [*Voss et al.*, 1984]. Compared to Early events, LEP events typically impact much larger regions of the ionosphere (sometimes >1000 km [*Peter and Inan*, 2004]). Since the increased ionization occurs over a larger region, two types of VLF perturbations, or a combination of both, is seen at the receiver [*Cotts*, 2011]. First, the disturbed ionospheric region may scatter some VLF energy towards the receiver, similar to Early event scattering. Secondly, if the ionosphere is significantly disturbed over a large (much greater than a wavelength) portion of the great-circle path, then the modal properties of the VLF wave change.

Figure 3.14 shows an LEP event measured by two separate receivers, Briarwood and Baxley, from a single transmitter, NLK. The causative stroke has a peak current of +204

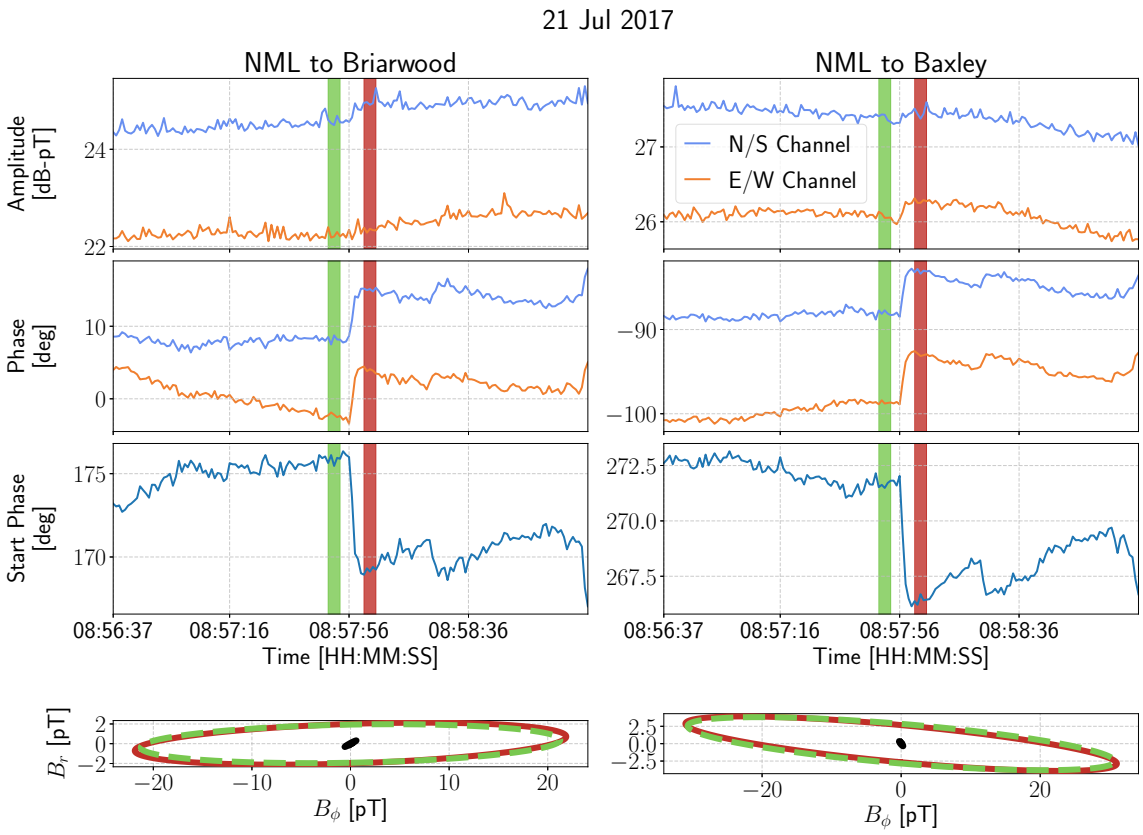

Figure 3.14: Polarization ellipse technique used on an LEP event, measured at two different receivers from a single transmitter. See the caption in Figure 3.13 for details on this figure layout.

kA and is located at 43.32°N and 101.17°W. The onset delay (the time difference from when the stroke occurs to when the perturbation is seen in the receiver data) is ~0.7 seconds, and the onset duration (period from when the perturbation occurs until the recovery begins) persists for over one second. Noise from other lightning strokes is particularly high during this event, so a 9-point median filter is applied to the low-resolution (1-Hz) data for visual clarity. The before and after ellipses from the NLK to Burden path show almost no change, and yet, a strong jump in the start phase occurred. This produced a very small scatter field ellipse shown at the origin as a small dot. For the other path, NLK to Baxley, a slight change in tilt angle and modification along the semi-minor axis occurred. Compared to the ellipse changes in the Early/fast event, the LEP event produced almost no change to the

shape of the polarization ellipse, just a change in start phase. The difference between this observation and the Early/fast event shown earlier may be due to the fact that Early/fast events result from scattering, whereas the LEP is caused by modal changes. In particular, the fact that the start phase changed whereas the shape did not may indicate that the phase velocity of the dominant propagating modes are impacted by the LEP event in a roughly equal way, but that the attenuation rate is not much affected. However, the effect of geometry of the transmitter, disturbance, and receiver, is also important and should not be discounted [*NaitAmor et al.*, 2010, 2013].

3.4.4 Solar Flare Event

The Sun is the chief driving force for the diurnal variations of the ionosphere. During the daytime, the Sun increases the electron density of the D region by orders of magnitude and creates a relatively stable and predictable ionosphere as the Sun traverses across the sky. LEP and Early/fast events are rarely, if ever, observed during the daytime. However, abnormally strong X-ray emissions from the Sun, known as solar flares, can perturb the daytime ionosphere an appreciable amount [*Thomson and Clilverd*, 2001]. Solar flares affect at least the entire daytime side of the ionosphere and enhance the electron density through the entire D region.

Figure 3.15 shows VLF perturbations from a C-class solar flare event over the continental US which is measured along the paths NLK to Burden and NLK to Baxley. The ellipses are median filtered over 4-minute windows. This example is interesting because the NLK to Burden narrowband clearly shows the solar flare in the amplitude data, but no change is visible in the phase data. In contrast, the NLK to Baxley path shows a disturbance plainly obvious in the phase, but barely detectable in the amplitude. This fact reinforces the idea that analyzing amplitude and narrowband data individually may not give a sufficiently comprehensive picture of the underlying phenomena and could either miss or mischaracterize ionospheric disturbances. Techniques such as scattered field method [*Dowden et al.*, 1996; *Gołkowski et al.*, 2014] and polarization ellipse method may provide crucial information in the analysis of these types of events.

The polarization ellipse shape before and after the solar flare ellipses are quite similar for

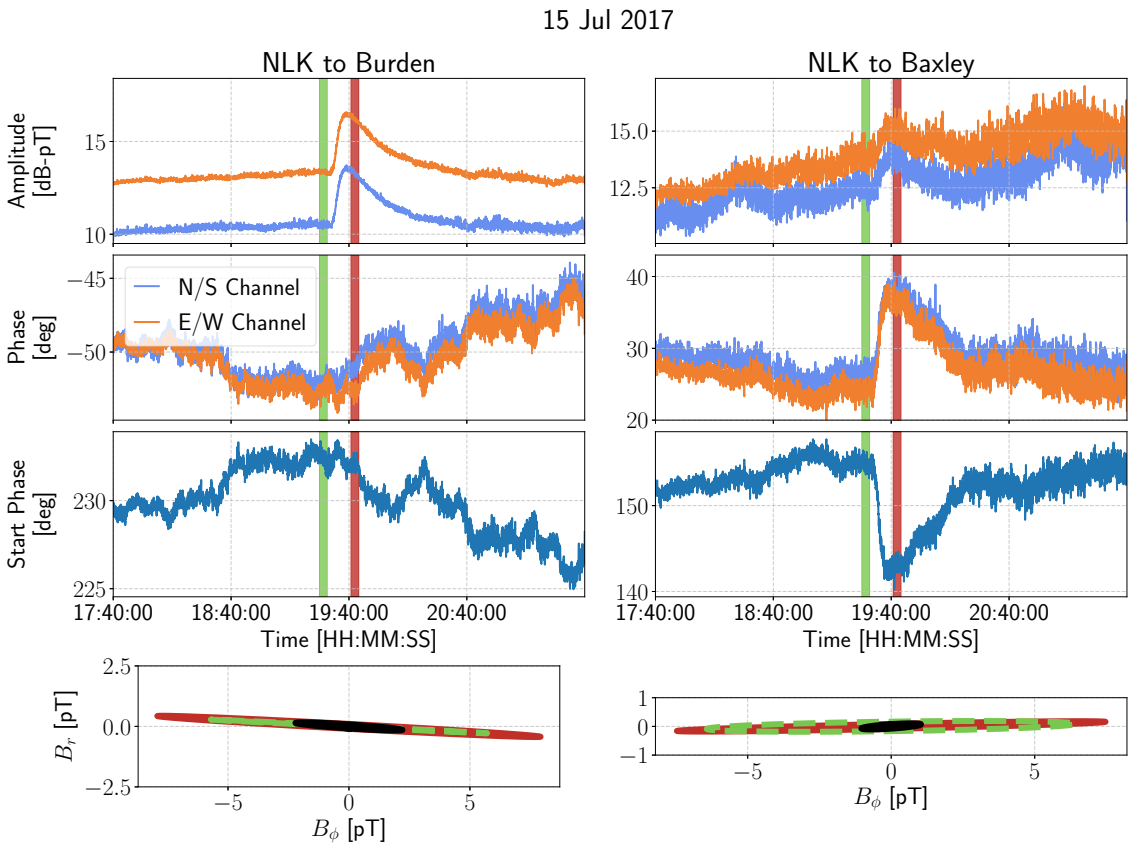

Figure 3.15: Polarization ellipse technique used on a C-class solar flare ionospheric disturbance, measured at two different receivers from a single transmitter. See the caption in Figure 3.13 for details on this figure layout.

both paths. Both experience minimal minor-axis change and no tilt angle change, but both present strong increases in the major-axis component. Likewise, the scattered field ellipses are almost linear and are close to parallel to the major axis of the before and after ellipses. The start phase on the NLK to Burden path shows almost no change from the solar flare, but the NLK to Baxley start phase has a strong change that is similar to the phase data but with opposite polarity. Since this event is during the daytime and over a long path, the transverse magnetic field dominates over the strongly attenuated radial magnetic field, which is why the ellipses are highly linear and almost no change occurs on the semi-minor component.

CHAPTER 4
COMPARISON OF VLF OBSERVATIONS TO LWPC

In this chapter, we examine the joint use of the LWPC program (from Chapter 2) and narrowband VLF processing tools (from Chapter 3) to infer the D region's electron density profile along a path of propagation. The methods used in this chapter are drawn from past works, and it is found that site dependent phase errors severely limit the flexibility of these methods.

4.1 LWPC Inverse Function Estimation

Most of the input parameters to the LWPC program are predefined by the chosen transmitter and receiver pair. Electrical properties of Earth's ground and the Earth's background magnetic field are considered static, and these values are simply found by querying a look up table, or some accurate model, for different points along the GCP. Likewise, the properties of the transmitter and receiver are also well established. The only variables we consider adjustable in the LWPC program are the waveguide parameters, h' and β. Waveguide parameters may be partitioned into discrete segments along the GCP, shown in Figure 2.4. Therefore, the GCP between a transmitter and receiver has a vector of waveguide parameters $\mathbf{h}'_n = [h'_1, h'_2, \ldots h'_\ell]^T_n \in \mathbb{R}^\ell$ and $\boldsymbol{\beta}_n = [\beta_1, \beta_2, \ldots, \beta_\ell]^T_n \in \mathbb{R}^\ell$, where there are ℓ segments along path n. Note that the length of vectors $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$ are path dependent and different paths may have different values for ℓ.

Modeling the azimuthal magnetic flux density $B_{\phi n} \in \mathbb{C}$ is done by inputting the waveguide parameter vectors into the LWPC program

$$L_n(\mathbf{h}'_n, \boldsymbol{\beta}_n) = B_{\phi n}, \tag{4.1}$$

where the function L_n represents the LWPC program with all the predefined parameters for path n. This function is backwards from the remote sensing problem – the magnetic

flux density is already known from radio receiver measurements, and the waveguide parameters are unknown because of the extreme difficulty in performing empirical D region measurements.

One thought for estimating the waveguide parameters is to use the inverse of L_n. An ideal inverse function will resolve the complete $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$ vectors. Unfortunately, neither L_n nor its inverse are one-to-one functions, so finding a unique inverse solution for L_n is not possible.

It has been shown by *Ferguson and Snyder* [1980] that the scalar mean, $\bar{h}'_n, \bar{\beta}_n \in \mathbb{R}$, of $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$ may be used as an approximation in LWPC to obtain similar results, meaning that the waveguide parameters can be assumed to be a constant value along the path, equal to the path-averaged waveguide parameters. This mathematically means

$$L_n(\mathbf{h}'_n, \boldsymbol{\beta}_n) \approx L_n(\bar{h}'_n, \bar{\beta}_n), \tag{4.2}$$

and it is valid under the condition that changes in electron density profile are small and slowly varying along the path. This approximation reduces the complexity of the inverse problem

$$L_n^{-1}(B_{\phi n}) \approx \left[\bar{h}'_n, \bar{\beta}_n\right] \tag{4.3}$$

by assuming the ionosphere is horizontally homogeneous along the entire path of propagation, and only the scalar values $\bar{h}'_n$ and $\bar{\beta}$ are solved for, instead of 2ℓ values (all elements of $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$).

4.2 Relating VLF Observations to Waveguide Properties

A complex magnetic field value is the output of the LWPC function L_n, but the amplitude of this value is found by assuming the power of the transmitter (this is one of the predefined values in the LWPC program), and the phase is referenced to some arbitrary time. In practice, the power of the transmitter is not precisely known. Transmitter power values are available in some of the VLF literature, but these values are usually presented with no citation, or the publication they cite has no citation. *Watt* [1967, p. 124] gives a comprehen-

sive table of 1960's VLF transmitter characteristics, but strongly cautions that some values may be incorrect due to changes in transmitter operation, upgrades, and inaccuracies from undisclosed sources used to produce the table. Operation of the NLK transmitter from 20 July 2017 to 28 August 2017 is one example of unexpected transmitter operation. During this period, the transmitter operated at a power level $\sim$ 10 dB lower than normal. This operation was not reported to the public and only a careful observer with historical data would have recognized a change in power. It is therefore advised to use published transmitter power values cautiously and to not rely on precise transmitter power for ionospheric remote sensing applications (unless a credible source with knowledge of the transmitter's current power is available).

Some methods do not require knowledge of the VLF source's power. *Han and Cummer* [2010a,b] used broadband radio sferics to infer waveguide parameters with a single receiver. Their approach involved spectral analysis of each sferic, and they used frequency dependent features to infer waveguide properties, all without needing to know the radiated VLF power from the lightning stroke. Narrowband VLF transmitters have a very small bandwidth (usually 100 kHz) making the spectral analysis method difficult to perform (results from *Koh et al.* [2018] underscore this difficulty). Spatial and temporal comparisons, rather than spectral, are well suited solutions to the transmitter power problem.

In the case of narrowband VLF transmitters, where transmissions are continuous at constant power for many hours, temporal comparisons are often performed with a single VLF receiver. The receiver will measure the complex magnetic or electric field over a large period of time, producing some variation in the measured field. This variation is the key to inferring the waveguide parameters. When using the temporal comparison method, VLF propagation is modeled with the LWPC program over a wide range of h' and β combinations, and the resulting normalized curve of the measured fields is compared with the modeled fields. Values for h' and β that give the best fit, when comparing LWPC to the observations, are chosen as the solution. *McRae and Thomson* [2000] present an excellent example of this method. One drawback of the temporal comparison method is it assumes constant power and phase of the transmitter. In Figure 3.11 it is shown that a VLF transmitter's phase is not guaranteed to be stable, likewise the level of radiated power from a transmitter is

not always constant. Thus, a change in transmitter operation may be misinterpreted as an ionospheric change, and incorrect for values for h' and β may be inferred.

Spatial comparison requires simultaneous operation of two or more VLF receivers spaced some distance apart. For just two receivers, the ratio of one receiver measurement to the other receiver measurement, taken at the same time, is compared with the LWPC program over a wide range of h' and β combinations. The advantage to this method is its independence of transmitter operation, meaning it is acceptable for the power or phase of the transmitter to change over time because the comparison is between receiver measurements taken at the same time. Geometry of the receivers is a concern when employing this method. Having the receivers too far apart may result in widely varying h' and β values between the two paths, which complicates the problem. It is best to have the two receivers close enough to each other so that the h' and β values are similar along both paths, but far enough apart so a reasonable change in the strength and phase of the field is measured. An ideal geometry would have two receivers with GCPs that lay on top of each other, but one receiver placed just a little closer to the transmitter. This would mean that h' and β, and all the other waveguide properties, are the same for both transmitter to receiver paths, except for the short offset between receivers. *Thomson* [2010] and *Thomson et al.* [2012] are examples of the spatial comparison method. Although, their methods rely on the assumption that the NWC transmitter has a radiated power of 1 MW.

4.2.1 Narrowband Phase Ambiguity Removal

Comparison of a narrowband signal's carrier phase, recorded at two receivers, requires the data acquisition system of each receiver to be synchronized with each other. The LF AWESOME receivers use GPS as a master clock to accomplish the timing synchronization. Demodulation of the MSK narrowband signal inherently includes a 90° ambiguity of the carrier phase ϕ_0 and clock phase ϕ_1, and this ambiguity must first be removed before carrier phase is compared between sites.

The mathematics for MSK demodulation (from equation (3.18)) reveal that the upper tone ($\phi_+(t) = \phi_0 + \phi_1 + m\pi$) and lower tone ($\phi_-(t) = \phi_0 - \phi_1 + n\pi$) have a 180° ambiguity ($m, n \in \mathbb{Z}$). The carrier phase equation $\phi_0 = [(\phi_+ - m\pi) - (\phi_- - n\pi)]$ / 2, from equation

(3.19), shows that any choice for m and n will adjust ϕ_0 by some multiple of 90°, and the same logic is also true for ϕ_1. Therefore, determining the values for m and n is the same as solving the 90° ambiguity problem.

A clock phase of 90° is equivalent to a time delay of 5 ms (for an MSK baud rate of 200 symbols per second), and 5 ms is the amount of time it takes a radio wave to travel ~1.5 Mm in free space. This means that if each receiver's distance from the transmitter is well known (much less than 1.5 Mm error range) then the 90° clock phase ambiguity is solved for by adjust the clock phase by multiples of 90° until the bit transition closely matches how long it would take a radio wave to propagate from the transmitter to the receiver. This process is equivalent to subtracting the time it took for the communicated signal to propagate from the transmitter to each receiver (group velocity) and then adjusting one of the signals by 5 ms increments until the two MSK signals closely match.

Removing the clock phase ambiguity constrains the carrier phase to a simpler 180° ambiguity. This is because the clock phase has an offset of $(m - n)\pi/2$, and the carrier phase has an offset of $(m+n)\pi/2$. If $m-n$ is an even number, then $m+n$ must also be even. It similarly follows for odd values of $m - n$. The 180° carrier phase ambiguity is removed by subtracting the synchronized clock phase and unsynchronized carrier phase from both trellises, and then comparing the phase trellises of the two signals and incrementing the offset of one phase trellis until the two signals match.

It is sometimes easier to understand how a problem is solved when real values are given. Consider two concurrent phase measurements of the NAA transmitter at the Briarwood and Baxley receivers, in Georgia, on 4 October 2017 at 19:00 UT. Briarwood is ~1.812 Mm from the transmitter, and Baxley has a distance of ~1.930 Mm, giving a differential distance of $\Delta d_{\mathrm{NAA}} = -118.0$ km. Briarwood measured a carrier phase of $\phi_{0,\mathrm{BW}} = -5.6°$ and a clock phase of $\phi_{1,\mathrm{BW}} = -78.8°$, and Baxley measured a carrier phase of $\phi_{0,\mathrm{BX}} = -68.0°$ and a clock phase of $\phi_{1,\mathrm{BX}} = 4.4°$. Clock phase is adjusted by finding the value of $m - n$ that makes $-\Delta d_{\mathrm{NAA}}$ / $[(\phi_{1,\mathrm{BW}} - \phi_{1,\mathrm{BX}} - (m - n)\ 90°)\ 4T\ /\ 360°]$ close to the speed of light (T is the symbol period). A value of $m - n = 1$ gives a value of $1.03c$ m/s, which is close to the speed of light, and is equivalent to a 5 ms adjustment to the symbol transition time τ_1. Any other choice for $m - n$ gives physically unreasonable results. Note that $1.03c$ is

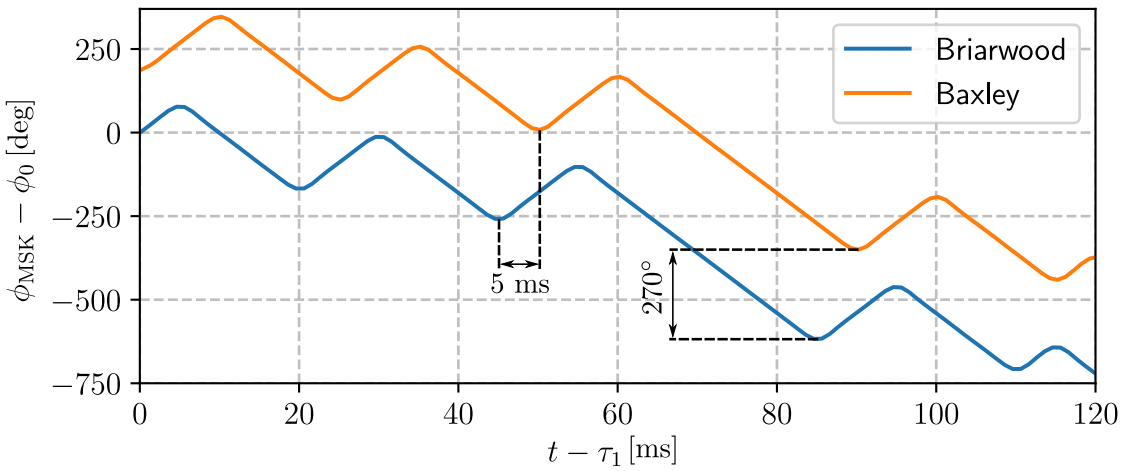

Figure 4.1: Phase trellis comparison between Briarwood and Baxley on 4 October 2017 at 19:00 UT. Baxley needs a 5 ms clock phase adjustment and 270° carrier phase adjustment to remove the 90° MSK demodulation ambiguity between Briarwood and Baxley.

not the group velocity since the wave is traveling along two different paths. However, if the two receivers were located on the same GCP and the signal consisted of just one mode, this would be an appropriate method for calculating group velocity along the GCP between the two receivers. In general signals are multi-modal so this result is an effective/aggregate group velocity.

In this example, $m - n$ is found to be an odd value, so a 90° adjustment must be made to the Baxley carrier phase, plus some multiple of 180°. The final piece of the problem is solved by looking at the in-phase and quadrature values (equivalent to looking at the phase trellis). In Figure 3.1, it is shown that the upper tone can have a phase of 0° or 180°, and similarly for the lower tone. By comparing the IQ data of both receivers, the phase of the tone is determined and adjusted if needed. In this example Briarwood has IQ values of -1 and 1, respectively, and Baxley has values of 1 and -1. Therefore, Baxley needs to be adjusted by 180° plus 90° (to account for the odd value of $m - n$), or $m + n = 3$, and m and n are found to be 2 and 1, respectively.

Carrier phase and clock phase corrections are also found through graphical means. Figure 4.1 shows the observed (not reconstructed) Briarwood and Baxley phase trellises ϕ_{MSK}. If the carrier and clock phase need no adjustment, then the phase trellis observed at the

Phase Wrapping p	Apparent Phase Velocity v_a [m/s]
8	$1.27c$
9	$1.12c$
10	$1.00c$
11	$0.91c$
12	$0.83c$

Table 4.1: Calculated phase velocity of the NAA signal given the number of spatial phase wrappings that occur when synchronizing the Briarwood and Baxley receivers.

two receivers should closely match when the uncorrected clock delay τ_1 is subtracted from time t, and the uncorrected carrier phase is subtracted from the observed phase trellis. The figure shows that Baxley lags Briarwood by 5 ms after the clock phase is subtracted from both signals. This 5 ms lag is equivalent to determining $m - n = 1$. Likewise, the Baxley phase trellis is offset from the Briarwood phase trellis by 270°, and this translates into $m + n = 3$.

At this point the phase demodulation ambiguity is removed, but the spatial phase wrapping, the number of 360° cycles p the wave makes along the path, still exists. It is known that the phase velocity of a VLF wave in the EIWG is close to the speed of light c, but the exact phase velocity is dependent on current EIWG properties. Phase velocity is not applicable in this case because Briarwood and Baxley have different GCPs to the NAA transmitter. The goal is to find the phase difference between the two receivers given the differential distance Δd_{NAA}. An apparent velocity value is determined from the phase difference and differential distance, but this velocity is not the same as phase velocity in the traditional sense; the term *apparent phase velocity* v_a is used to avoid confusion with the traditional phase velocity definition.

The NAA transmitter (24.0 kHz) has a free space wavelength λ of ~12.5 km, and determining the spatial phase wrapping for this example is not difficult, since $\Delta d_{\mathrm{NAA}} \approx 9.5\lambda$. Table 4.1 gives the calculated apparent phase velocity for different 360° phase wrapping values. LWPC reveals that the only reasonable value for p is 10 (the modeled apparent phase velocity from LWPC is roughly between $1.00c$ and $1.02c$, but results are highly dependent on waveguide parameters and GCP geometry). This gives a final carrier phase differential of $\Delta\phi_0 = \phi_{0,\mathrm{BX}} + 270° - \phi_{0,\mathrm{BW}} - 360°\ p = -3392.4°$.

Phase Wrapping p	Apparent Phase Velocity v_a [m/s]
99	$1.02c$
100	$1.01c$
101	$1.00c$
102	$0.99c$
103	$0.98c$

Table 4.2: Calculated phase velocity of the NAA signal given the number of spatial phase wrappings that occur when synchronizing Briarwood and Lost Pines receivers.

The relatively small differential distance between the Briarwood and Baxley receivers makes finding the apparent phase velocity simple, but the task becomes more difficult when the two receivers are far apart. Consider the same differential phase problem but replace the Baxley receiver with the Lost Pines receiver in Texas. This gives a differential distance of $\Delta d_{\text{NAA}} = -1.26$ Mm. Table 4.2 presents a list of apparent phase velocities, which range from $0.98c$ to $1.02c$, all of which are similar to what LWPC predicts. The problem is further complicated because h' and β may be quite different over the two paths, and it can no longer be assumed that the apparent phase velocity is close to the speed of light. It is therefore not possible to solve the phase wrapping problem between the Briarwood and Lost Pines receivers, unless more information is included, such as using more receivers that are located between the Briarwood and Lost Pines receivers.

4.2.2 A Simple Approach to Inferring Waveguide Parameters

Now that the spatial and demodulation ambiguities have been removed from the Briarwood and Baxley phase differential measurement, the spatial comparison method is used to determine the average $\bar{h}'$ and $\bar{\beta}$ values along the two similar GCPs. The fundamental assumption for this method is that $\bar{h}'$ and $\bar{\beta}$ are the same for both paths. This assumption is valid since the paths are close together and the time being considered is near minimum solar zenith (mid-day), but this assumption is most likely not valid during the nighttime, when the terminator is near the GCPs, or when the paths are far apart.

Finding $\bar{h}'$ and $\bar{\beta}$ through the spatial comparison method requires a modification of the LWPC inverse function, $L^{-1}(B_{\phi n}) = \left[\bar{h}', \bar{\beta}\right]$, but how to find the inverse function has not been discussed. The simplest implementation of the spatial comparison method is the

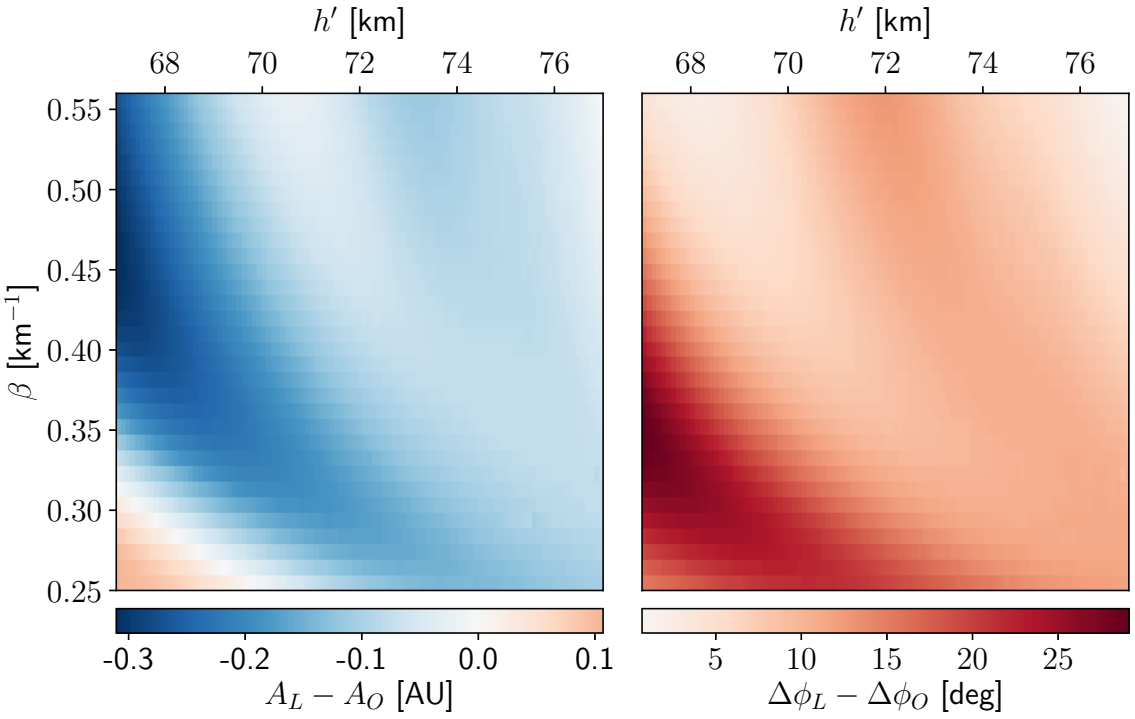

Figure 4.2: Differential measurement between Briarwood and Baxley compared with LWPC.

brute force approach – run LWPC over a range of h' and β values and choose the h' and β combination that minimizes

$$\left[\bar{h}', \bar{\beta}\right] = \arg\min_{h',\beta} \left\{ \left| \frac{L_{\mathrm{BW}}(h', \beta)}{L_{\mathrm{BX}}(h', \beta)} - \frac{B_{\phi\mathrm{BW}}}{B_{\phi\mathrm{BX}}} \right| \right\}. \tag{4.4}$$

Note that the complex ratio $B_{\phi\mathrm{BW}}$ / $B_{\phi\mathrm{BX}}$ is independent of any change in transmitter power and any transmitter phase instabilities, and the determined $\bar{h}'$ and $\bar{\beta}$ values are indifferent to transmitter operation.

Figure 4.2 shows the error between the modeled and observed complex magnetic flux density for a range of h' ($67 - 77$ km) and β ($0.25 - 0.56$ km^{-1}) values. The panel on the left is the difference between the modeled ratio amplitude $A_L = |L_{\mathrm{BW}}(\bar{h}', \bar{\beta}) \ / \ L_{\mathrm{BX}}(\bar{h}', \bar{\beta})|$ and observed ratio amplitude $A_O = |B_{\phi\mathrm{BW}} \ / \ B_{\phi\mathrm{BX}}|$, and the panel on the right is the difference between the modeled ratio angle $\Delta\phi_L = \angle\left[L_{\mathrm{BW}}(\bar{h}', \bar{\beta}) \ / \ L_{\mathrm{BX}}(\bar{h}', \bar{\beta})\right]$ and observed ratio angle $\Delta\phi_O = \angle[B_{\phi\mathrm{BW}} \ / \ B_{\phi\mathrm{BX}}]$. The solution to (4.4) is $\bar{h}' = 76.8$ km and $\bar{\beta} = 0.5$ km^{-1}. This solution does not match well with past findings [*Morfitt*, 1977; *Ferguson*, 1992;

McRae and Thomson, 2004; *Thomson*, 2010; *Thomson et al.*, 2012]. The $\bar{\beta}$ value is not unreasonable, but the $\bar{h}'$ value is much too large.

Interpretation of the results from *Thomson et al.* [2012] shows $A_L - A_0$ and $\Delta\phi_L - \Delta\phi_0$ will transition from negative to positive (or vice versa) as the correct set of $\bar{h}'$ and $\bar{\beta}$ is approached. This transition is not seen in Figure 4.2. Instead $A_L - A_0$ is almost always negative and $\Delta\phi_L - \Delta\phi_0$ always positive, and this gives little confidence that the estimate of $\bar{h}' = 76.8$ km and $\bar{\beta} = 0.5$ km^{-1} is correct.

4.3 Radio Wave Scattering

4.3.1 Briarwood Differential Phase Experiment

Unexpected results of the spatial comparison method (Figure 4.2) prompted an investigation into the Briarwood receiver's performance. A series of differential phase measurements were taken near the Briarwood receiver with an identical LF AWESOME receiver. These differential measurements reveal that near field scattering, from nearby conductive objects, has a non-negligible impact on narrowband VLF phase measurements, which can exceed 10° under certain receiver and scattering object geometries. Chapter 3 presents a quadrantal error correction method, which is intended to mitigate the impact of angle of arrival errors from nearby scatters. While this angle of arrival correction does affect measured phase slightly, its impact is generally much less than what is measured during this investigation, and it is concluded that nearby radio wave scattering is not completely mitigated by quadrantal error corrections.

Results for the differential phase measurements near the Briarwood receiver are shown in Table 4.3, and the location for each differential measurement is shown in Figure 4.3. Carrier phase measurements of the NAA transmitter are taken at the stationary Briarwood receiver and portable receiver, and the difference between these two measurements is $\Delta\phi$. The portable receiver is positioned in different locations along the NAA GCP, defined by the NAA transmitter and Briarwood receiver. Differential distance Δd_{NAA} is calculated as the distance from NAA to the Briarwood receiver less the distance from NAA to the portable receiver. A free space differential phase $\Delta\theta$ is also calculated to give a reference

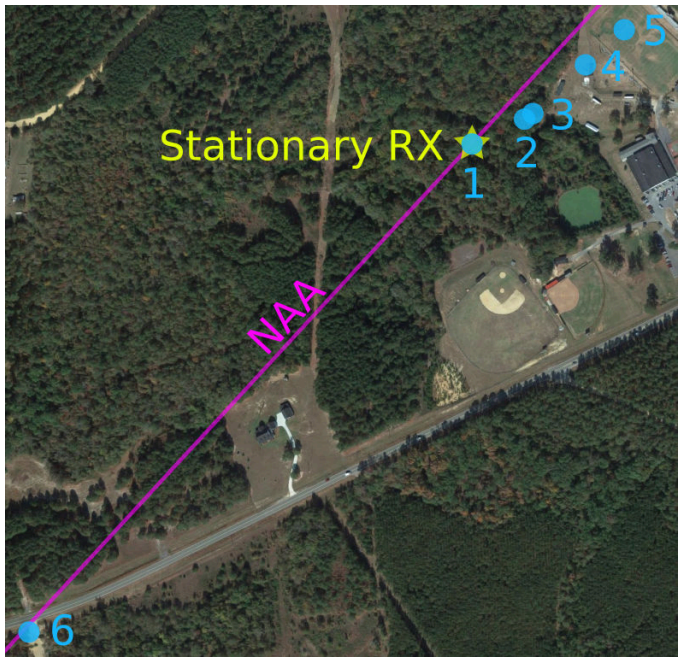

Figure 4.3: Map for the Briarwood differential phase measurements. The Briarwood receiver is kept stationary (yellow star) and a mobile receiver (blue points $1-6$) is moved to different locations near the NAA GCP (purple line).

as to what the approximate differential phase measurement should be.

For the first measurement, the antenna of the portable receiver is positioned at the same location as the Briarwood antenna. This measurement is used to calibrate the differences between the receivers, most notably, cable length. The exact cable length of the Briarwood receiver is unknown, but it is estimated (this estimation is based on the recollection of those involved in the installation of the Briarwood receiver) to be about 244 m to 305 m (800 ft to 1000 ft), and the portable receiver has a cable length of 8 m. With the two antennas in the same location, the differential distance is zero and a differential carrier phase of $-14.6°$ is measured. This $-14.6°$ phase difference is mostly attributed to the Briarwood receiver's extra cable, and it is used to calibrate the differential phase measurements between the two receivers (i.e., $-14.6°$ is subtracted from all $\Delta\phi$ measurements to account for the difference in cable length). Imperfect tolerances of a receiver's electronics also cause small phase discrepancies between receivers, but in this case signal cable length is the dominant mechanism for carrier phase disparity.

All LF AWESOME receivers operated by the Georgia Tech LF Radio Lab use a Belden

Number	Δd_{NAA} [m]	$\Delta\theta$ [deg]	$\Delta\phi$ [deg]	$\Delta\theta - \Delta\phi$ [deg]	v_p [m/s]
1	0	0.0	0.0	0.0	$0.00c$
2	65	-1.9	1.0	-2.9	$-1.87c$
3	71	-2.0	1.6	-3.6	$-1.28c$
4	163	-4.7	-4.5	-0.2	$1.04c$
5	222	-6.4	-16.3	9.9	$0.39c$
6	-748	21.6	37.0	-15.4	$0.58c$
7	0	0.0	0.4	-0.4	$0.00c$

Table 4.3: Results for the Briarwood differential phase test on 5 June 2018. The difference in GCP path length for the two receivers is Δd_{NAA}, the difference in measured NAA carrier phase is $\Delta\phi$, and the inferred phase velocity is v_p. A phase difference for free space propagation over distance Δd_{NAA} is given as $\Delta\theta$.

1217B 22AWG 4 pair low-capacitance cable with a capacitance of $C' = 62.339$ pF/m and an inductance of $L' = 0.6562$ μH/m. From transmission line theory, the phase velocity of the signal cable is found to be $v_p = 1/\sqrt{L'C'} = 1.56 \cdot 10^8$ m/s, or $0.52c$ m/s. Using the cable's phase velocity and the measured $-14.6°$ differential carrier phase, a length of ~ 271 m is calculated for the Briarwood signal cable, which agrees with the estimated cable length of 244 m to 305 m. A final measurement, number 7 in the table, is performed to ensure the receivers are still calibrated after all other measurements have been taken. With the mobile antenna placed next to the Briarwood antenna, in an identical position as the first measurement, a $\Delta\phi = 0.4°$ is measured. This gives confidence in the alignment of the GPS synchronized sampling clocks, and that the relative phase calibration of $-14.6°$ is correct.

Measurements 2 through 6 are located along the NAA to Briarwood GCP. This is done to indirectly measure phase velocity v_p, which is only possible when the two receivers reside on the same GCP. Results for the inferred phase velocity reveal an issue with this method; measurement 3 has a phase velocity of $-1.28c$ and then measurement 4, just a 92 m difference, has a phase velocity of $+1.04c$. Having the mobile receiver so close to the stationary receiver is one reason for the large discrepancies in phase velocity. The smaller Δd_{NAA} is, the more sensitive the phase velocity calculation becomes to noise in the $\Delta\phi$ estimate. Imperfect carrier phase estimates may be a small piece of the problem, but the errors in the inferred phase velocity are much too large to be caused by this alone.

Briarwood is located in a rural and sparsely populated region of Georgia, but still, human infrastructure in this area may be an issue. Nearby the Briarwood receiver are

power lines, underground pipes and cables, playground equipment, and a football field with a fence and tall towers for flood lights. All of these objects are conductive, and scattering of radio waves from these conductive objectives may be the major culprit for the discrepancies between the differential phase measurements and what waveguide theory predicts. Given the number of potential scattering objects nearby, it is difficult to determine if any or all the objects are causing the differential phase errors. Thus, a more controlled environment is needed to isolate the radio wave scattering impact on differential phase measurements.

Past works have touched on the nearby radio wave scattering problem, but their mention of the issue is often brief and in passing. The most common discussion for site errors involves angle of arrival errors for lightning location systems [e.g., *Mach et al.*, 1986; *Wood and Inan*, 2004; *Zoghzoghy*, 2015]. Angle of arrival errors are mitigated using the angle of arrival correction method in Chapter 3, but this method does not totally correct carrier phase errors of narrowband signals. *Bainbridge and Inan* [2003] present differential phase measurements, for an array of receivers, and indirectly discuss the radio wave scattering problem. Their goal is to remove any phase bias from radio wave scattering so that they can decompose the properties of individual waveguide modes. *Bainbridge and Inan* [2003], however, stop short of giving an in-depth discussion on the causative mechanism for the phase biases, and they only state that the biases need to be removed for the analysis. *Thomson* [2010] reports that one receiver suffered from radio wave scattering interference from buried conductors. Unfortunately, they never discuss how the buried conductor is determined to be the culprit, or how they concluded there is an interference problem at all. *Füllekrug et al.* [2014] uses 10 radio receivers within a 1 km $\times$ 1 km area to determine the slowness and azimuth arrival of the wave number vectors associated with narrowband VLF radio waves. *Füllekrug et al.* [2014] does not report on any site error effects or calibration, but the array does have the potential to be a useful tool in studying radio wave scattering and its effects on carrier phase.

4.3.2 Scherer Differential Phase Experiment

Topological features like mountains, changes in ground conductivity, and human infrastructure make finding an environment suitable for controlled VLF measurements difficult

Number	Δd_{NAA} [m]	d_{TL} [m]	$\Delta\theta$ [deg]	$\Delta\phi$ [deg]	$\Delta\theta - \Delta\phi$ [deg]
1	0	942	0.0	0.0	0.0
2	-33	908	1.0	2.3	-1.3
3	-79	856	2.3	1.8	0.5
4	-130	790	3.7	5.2	-1.5
5	-198	699	5.7	7.6	-1.9
6	-369	516	10.6	8.6	2.0
7	-604	417	17.4	14.2	3.2
8	-747	211	21.5	21.6	-0.1
9	-749	76	21.6	28.4	-6.8
10	-731	0	21.1	20.1	1.0
11	-738	-50	21.3	9.7	11.6
12	-799	-131	23.0	15.2	7.8
13	-953	-322	27.5	33.7	-6.2

Table 4.4: Results for the Scherer differential phase test on 19 June 2018. The difference in GCP path length for the two receivers is Δd_{NAA}, the difference in measured NAA carrier phase is $\Delta\phi$, and the distance of the portable receiver to the nearest point of the transmission lines is d_{TL}. A phase difference for free space propagation over distance Δd_{NAA} is given as $\Delta\theta$.

[*Rakov*, 2016, p. 165]. The ideal location for studying the impact of radio wave scattering should have a single conductive object for scattering radio waves and no other conductive objects within the vicinity.

A location (33.071°N, 83.834°W) near the Scherer Power Plant, operated by Georgia Power, is found to have the desired qualities for a radio wave scattering investigation. This site is close enough to Georgia Tech for a day-trip, and it is isolated from human infrastructure. The only identifiable conductive objects are a single row of high voltage transmission lines coming from the power plant. No buildings, buried conductive material, or other transmission lines are found or known near the site. These transmission lines are considered to be the sole scattering object in this controlled environment, and it is hypothesized that this scattering object will induce anomalous results in the differential phase measurement when a receiver is placed near the object.

Results for the Scherer differential phase experiment are shown in Table 4.4. The same notation from the Briarwood scattering experiment is used here, and one extra term is included, transmission line distance d_{TL}, which is the distance of the portable receiver to the closest point of the transmission lines. The two receivers have signal cables with similar

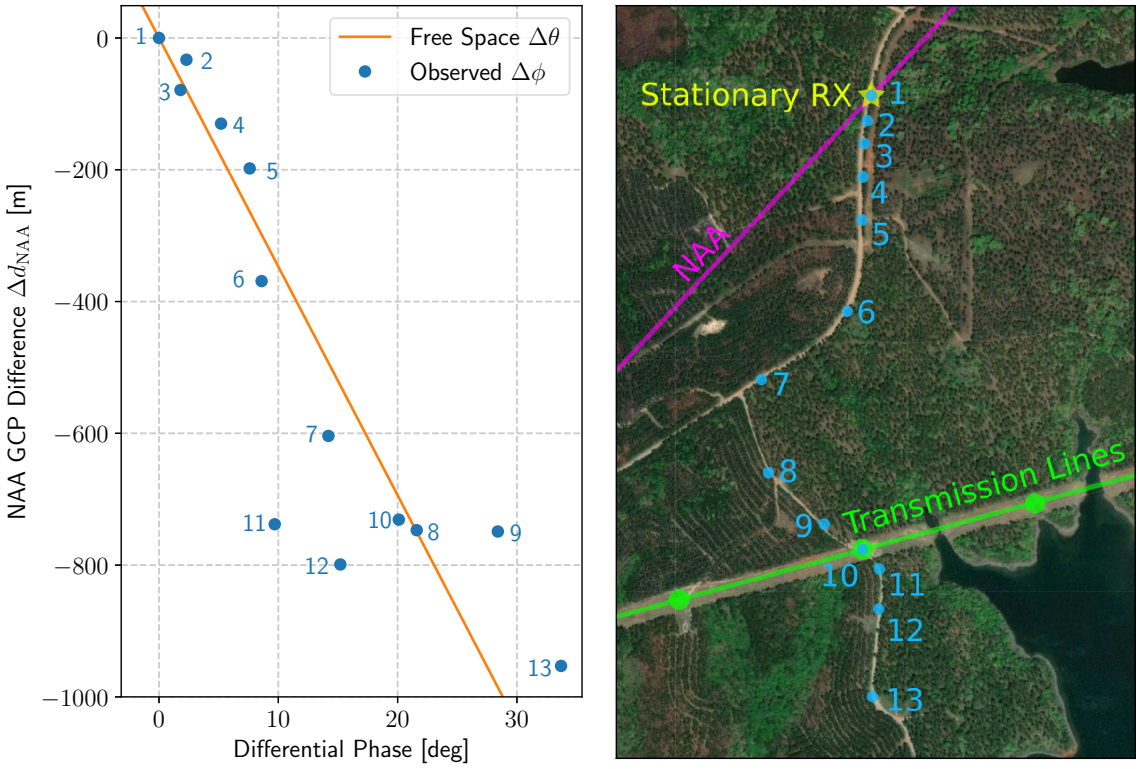

Figure 4.4: Differential phase $\Delta\phi$, of the NAA transmitter (purple line is NAA's GCP), between the stationary receiver (yellow star) and a mobile receiver (blue points $1-13$). Δd_{NAA} is the GCP distance from NAA to the stationary receiver less the GCP distance from NAA to the mobile receiver. The transmission lines (green line) and towers (green circles) are the near field radio wave scattering objects.

lengths, and a calibration differential phase value of $-2.4°$ is found. While moving the receiver from location number 7 to location number 8, the NS lowpass filter card disconnected from the motherboard. Therefore, measurement numbers 8 through 13 consist of only EW antenna measurements. Loss of the NS channel slightly affects the differential phase measurement by adding a small bias, but this small bias has little impact on the analysis used here.

Figure 4.4 shows the results of the differential phase measurements in the left panel and the location of receiver measurements in the right panel. A stationary receiver, marked with a yellow star, is placed far from the transmission lines, and a second mobile receiver is moved to thirteen different locations, marked with blue circles. When the receiver is far from the transmission lines, shown as a green line (green circles on the line are where the

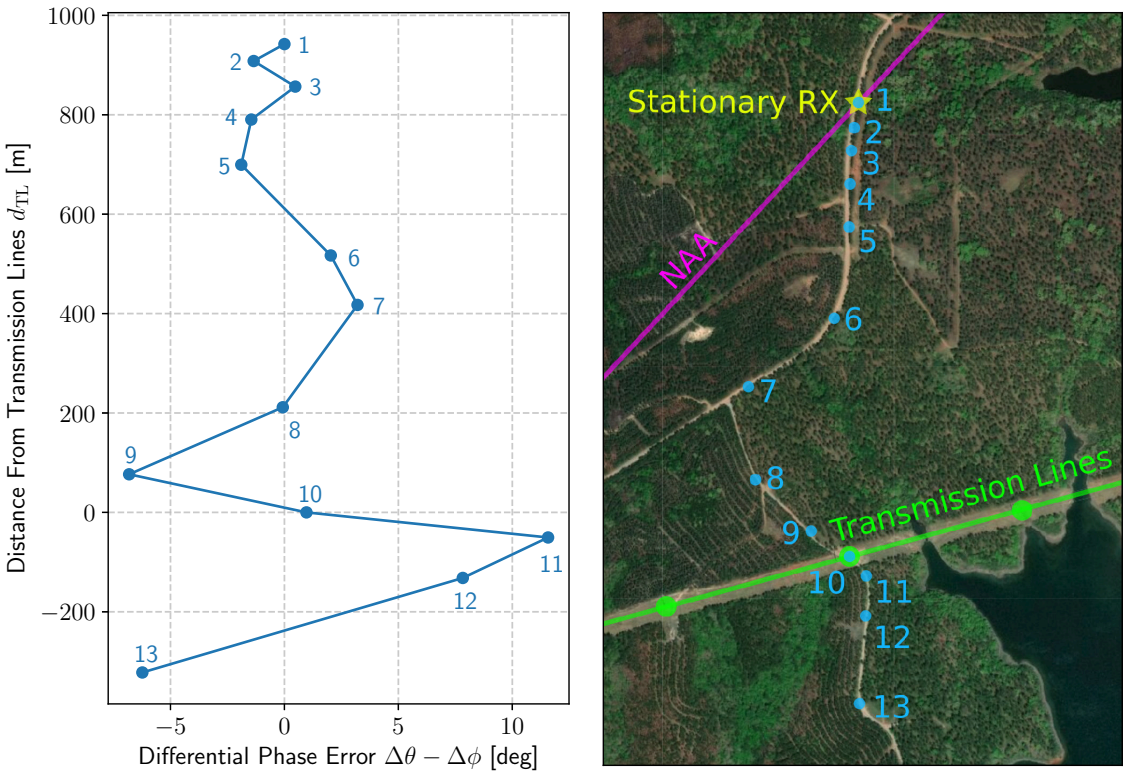

Figure 4.5: Differential phase error $\Delta\theta - \Delta\phi$, of the NAA transmitter (purple line is NAA's GCP), between the stationary receiver (yellow star) and a mobile receiver (blue points $1 - 13$). d_{TL} is the shortest distance of the mobile receiver to the transmission lines. The transmission lines (green line) and towers (green circles) are the near field radio wave scattering objects.

towers that support the cables are located), the differential phase $\Delta\phi$ seems to be relatively stable and close to the free space differential phase $\Delta\theta$. The differential phase measurements become more erratic as the portable receiver approaches the transmission lines, but it is difficult to see a pattern in the $\Delta\phi$ measurements.

A differential phase error $\Delta\theta - \Delta\phi$ is found by assuming $\Delta\phi$ should be the free space differential phase $\Delta\theta$. This assumption is not exactly true and contains about a 2% error (according to LWPC modeling), but it is a good assumption for this analysis. Figure 4.5 shows the differential phase error, and the distance from the transmission lines d_{TL} is used instead of Δd_{NAA}. Displaying the measurements in this manner reveals a clear oscillation that is strongest when the mobile receiver is near the transmission lines. An interesting point is that the measurement directly under the transmission lines, number 10, has a small

differential phase error, but measurement numbers 9 and 11 have the largest errors with opposite polarity.

Radio wave scattering is a problem for both spatial comparison (comparing observations between two receivers taken at the same time) and temporal comparison (comparing observations from the same receiver at different times). When everything is stationary, radio wave scattering off nearby objects imposes a constant bias on the observed phase and amplitude. If the transmitter were to then change its phase or amplitude by some amount, that change in phase or amplitude would be the same with or without the radio wave scattering effect (i.e., the change is independent of the bias for stationary systems). But if the waveguide were to change, meaning a change in h' and β, the bias from the radio wave scattering will not generally remain constant. Changes in the waveguide cause changes in the propagating modes, and a change in the modes (amplitude or phase) will influence the bias induced by the radio wave scattering. The radio scattering bias is more sensitive to ionospheric changes when one of the dominant modes is near a null or when there are many dominant modes in the waveguide.

It is unfortunate the NS channel was lost after measurement number 7. Scatter ellipse analysis from Chapter 3 may give deeper insight towards this radio scattering problem. Nevertheless, this experiment, in a controlled VLF environment, shows radio scattering is a concern when conductive objectives are within the vicinity of the receiver, and it can have a profound effect on carrier phase estimates.

To conclude this chapter, we want consider effects nearby radio wave scattering may have on h' and β inference methods. Recall the $\Delta\phi_L - \Delta\phi_O$ error in Figure 4.2. From the results of the Scherer phase experiment, it is not unreasonable to say a 10° error may exist in the $\Delta\phi_O$ estimate. If we apply a 10° site error adjustment to the $\Delta\phi_O$ estimate ($\Delta\phi_L - \Delta\phi_O - 10°$), then the results would closely match those from *Thomson et al.* [2012]. This does not mean 10° is the correct calibration value, or that site errors are definitively the problem with the $\Delta\phi_L - \Delta\phi_O$ error, but it does suggest radio wave scattering is a logical and plausible culprit for the discrepancy between VLF observations and LWPC.

CHAPTER 5
CONCURRENT ELECTRON DENSITY INFERENCE

Under normal daytime conditions, GCPs that are very similar to each other typically have similar electron density profiles. In chapter 4, we present a method that solves for the waveguide parameters along two similar paths, but the method is limited because it assumes both paths have the same $\bar{h}'$ and $\bar{\beta}$ values. Since our goal is to eventually infer waveguide parameters across a network of paths that are very different from each other, this assumption cannot be maintained. In this chapter, we will relax this constraint and allow different GCPs to have different waveguide parameters. We develop a scalable inverse function using machine learning that uses a general number of transmitter to receiver observations to concurrently solve for the waveguide properties along all paths.

5.1 Artificial Neural Network

5.1.1 A LWPC Based Inverse Function for Multiple Paths

Modeling the magnetic flux density B_ϕ is done with LWPC when $\bar{h}'$ and $\bar{\beta}$ are known. For N transmitter to receiver paths, the waveguide parameter vectors are $\bar{\mathbf{h}}', \bar{\boldsymbol{\beta}} \in \mathbb{R}^N$ and the magnetic flux density vector is $\mathbf{B}_\phi \in \mathbb{C}^N$. A function for these N LWPC runs is compactly written as

$$F(\bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}) = \begin{bmatrix} L_1(\bar{h}'_1, \bar{\beta}_1) \\ \vdots \\ L_N(\bar{h}'_N, \bar{\beta}_N) \end{bmatrix} = \begin{bmatrix} B_{\phi 1} \\ \vdots \\ B_{\phi N} \end{bmatrix} = \mathbf{B}_\phi, \tag{5.1}$$

and the inverse of F is thus

$$F^{-1}(\mathbf{B}_\phi) = \begin{bmatrix} \bar{h}'_1 & \bar{\beta}_1 \\ \vdots & \vdots \\ \bar{h}'_N & \bar{\beta}_N \end{bmatrix} = \left[\bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}\right]. \tag{5.2}$$

It is important to note that finding the magnetic flux density $B_{\phi n}$ on path n is only dependent on the waveguide parameters for that path, $\bar{h}'_n$ and $\bar{\beta}_n$, and all other elements in $\bar{\mathbf{h}}'$ and $\bar{\boldsymbol{\beta}}$ have no impact on determining $B_{\phi n}$. Conversely, finding $\bar{h}'_n$ and $\bar{\beta}_n$, with function F^{-1}, does not solely depend on $B_{\phi n}$, but it instead generally depends on all elements of $\mathbf{B}_\phi$. This property for F^{-1} satisfies the conditions for the new inverse function being sought: the elements of $\bar{\mathbf{h}}'$ and $\bar{\boldsymbol{\beta}}$ are not required to be the same, and magnetic flux density observations $\mathbf{B}_\phi$ along all paths are used to concurrently infer $\bar{\mathbf{h}}'$ and $\bar{\boldsymbol{\beta}}$.

To help constrain the estimation of F^{-1}, the GCPs are assumed to be close enough to each other so that $\bar{h}'$ and $\bar{\beta}$ values are partially correlated between paths. This assumption is valid when the ionosphere is smooth and undisturbed under all paths, or when only large-scale perturbations exists in the ionosphere. A large-scale perturbation is defined as a region that is smoothly perturbed and covers all, or close to all, the paths. During the daytime, ionization of the ionosphere is stable and varies slowly both spatially and temporally. For example, solar flares, which perturb the entire daytime ionosphere, satisfy this large-scale assumption.

In practice, function F is easy to calculate because it consists of multiple runs of the forward model, LWPC. Finding F^{-1}, on the other hand, is not easy. One approach to finding F^{-1} is to derive an inverse model, but this would be much more of a challenge than developing the LWPC forward model. Another approach would be to somehow estimate F^{-1} while still using LWPC as the backbone for the propagation physics. Estimation of F^{-1} would require some assumption about the function itself, such as assuming F^{-1} is a linear system, but a linear system assumption would be a poor choice for this type of problem. There is no question that the LWPC function L is a non-linear function, and its inverse L^{-1} does not generally have a unique solution. Thus, some non-linear function estimation method should be used to find F^{-1}.

Artificial neural networks (ANNs) have been shown to be good approximators for non-linear functions [*Haykin*, 1999, p. 230]. ANNs are a common subset of machine learning methods, and their incredible flexibility has made them useful for all sorts of industrial and research applications, such as image recognition, forecasting, and feedback controllers. In the simplest sense, an ANN may be thought of as a function with inputs and outputs that

take on numeric values.

One reason for the popularity of ANNs is their ability to estimate complicated non-linear functions. In fact, ANNs with just one single hidden layer (what a hidden layer is will soon be discussed) have been shown to be universal approximators [*Hornik et al.*, 1989], which means an ANN can approximate a continuous function given the appropriate construction of the ANN. Even though ANNs have this approximation power, it does not mean an ANN is always the optimal choice for estimating a function given the complexity required, the training time, and its generalization.

5.1.2 ANN Basics

There are many uses for ANNs, but the three common categories are classification, clustering, and regression. Classification is the task of predicting categorical variables. An example of classification is to classify what type of fruit is in a image (e.g., apple, orange, banana). Clustering is the act of forming natural groupings between observations. Clustering is great for grouping together multidimensional observations in ways that are foreign to human intuition. The final category, and the one we are interested in, is regression. Regression is the method of inputting some information into an ANN and then receiving continuous numerical outputs. Regression is the obvious choice for this work because the inputs and outputs to function F^{-1} are numeric values.

Many flavors of neural networks have been formulated to solve different machine learning problems. Here we utilize one of the simplest forms of ANNs, a feed-forward artificial neural network (which is referred to plainly as an ANN from here on). The following will emphasize how an ANN can be used to estimate our inverse function F^{-1}, and only a brief background of ANN fundamentals will be given. Some suggested books for ANN and general machine learning theory: *Haykin* [1999] gives a full overview of ANNs, *Harrington* [2012] is a nice beginner's introduction to machine learning concepts, and *Hastie et al.* [2001] covers more advanced machine learning theory.

Figure 5.1 is a graph for a simple ANN. An input vector $\mathbf{x} \in \mathbb{R}^2$ is shown on the left side and an output vector $\mathbf{y} \in \mathbb{R}$ is shown on the right. Arrows connect the input and output vectors to *nodes* (the circles in the graph), and each column of nodes is referred to

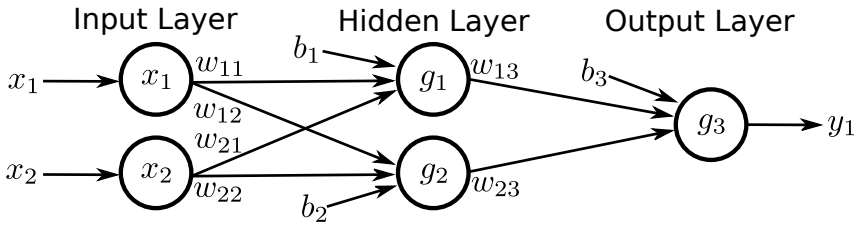

Figure 5.1: A simple ANN graph with input, hidden, and output layers.

as a layer. Each node in a layer is connected to all nodes in the previous layer (one layer to the left), excluding the input layer. An ANN may have multiple hidden layers, but for simplicity the graph is shown to only have one. This design is motivated somewhat by the layout of an animal's brain; neurons in the ANN are inspired by a brain's neurons, and the arrows connecting the ANN's neurons mimic a brain's synapses. In the graph, all the arrows point from left to right, which is why this neural network version is referred to as a feed forward neural network; information only flows in the forward direction during normal operation.

Mathematically, the neural network is defined by two main components: the activation functions $g(\cdot)$ and the weights $w, b \in \mathbb{R}$. It is easiest to see the relationship between vectors $\mathbf{x}$ and $\mathbf{y}$ by working backwards through the ANN. The output value y_1 is the output of activation function $g_3(z_3)$, and the argument z_3 of the activation function is the sum of the three values pointing to the node in the output layer

$$z_3 = b_3 + \sum_{i=1}^{2} w_{i3}\, g_i(z_i)\ . \tag{5.3}$$

Input values to activation functions $g_1(z_1)$ and $g_2(z_2)$ are found in a similar way

$$z_k = b_k + \sum_{i=1}^{2} w_{ik} x_i\ ,\ k = 1, 2. \tag{5.4}$$

Non-linear behavior of the ANN comes from the activation functions. There are many different activation functions to choose from, and the choice of which one to use is problem dependent. A widely used activation functions is the sigmoid function $g(z) = 1/\left(1 + e^{-z}\right)$. This function is non-linear and has a somewhat similar shape to the Heaviside step function,

except that the step function has an abrupt transition at $z = 0$ whereas the sigmoid function has a smooth transition in its output near $z = 0$. Another non-linear activation function that is gaining popularity is the rectified linear unit (ReLU) $g(z) = \max(0, z)$. Both the sigmoid function and the ReLU have a lower bound of zero, but the ReLU upper bound is infinite. Activation functions can also be linear $g(z) = \alpha z$, where α is usually unity. It is standard practice for ANN regression problems to use linear activation functions with unity slope for the output nodes. If all the activation functions in the ANN are linear functions, then the ANN reduces to a linear system. This work uses ReLU activation functions for the hidden layer nodes (it is found through ANN tuning that ReLU's give good performance for the problem presented in this thesis), and linear activation functions with unity slope are used for the output nodes.

Layers, nodes, and activation functions form the mathematical framework for the ANN. Once these are chosen, the statistical performance of the ANN resides in the determination of the weights. A set of $\mathbf{x}$ and $\mathbf{y}$ vectors, known as training data, are used in a supervised learning method to statistically find weights that give the best ANN performance. How large this set of training data needs to be depends on the problem at hand. If the ANN is expected to perform a complicated task (the target function) then it will most likely have a larger number of neurons and possibly layers. Adding more neurons or layers increases the number of weights in the network, which gives the ANN more flexibility in approximating the target function. Choosing enough neurons and layers reduces the bias error (how much of an assumption is made about the target function). Having enough unique samples in the training data that spans the scope of the problem is also important. A high level of variance (changes in the estimate of the target function when different training data sets are used) will ensue if the training data does not properly represent the entire domain and range of the problem. Thus, to minimize variance and bias, the ANN must be constructed with enough complexity (layers, nodes, and activation functions), and training data must give good coverage of the problem at hand.

Training the neural network (adjusting the weights) is done through a method known as backpropagation. How well the neural network estimates the target function is found by applying a single training data sample $\mathbf{x}$ to the ANN and comparing the output of the ANN

$\mathbf{y}$ to the target (truth) vector from the training data $\mathbf{y}'$ through a loss function $E(\mathbf{y}, \mathbf{y}')$. The loss function we use in this work is the squared error

$$E(\mathbf{y}, \mathbf{y}') = \frac{1}{2}||\mathbf{y} - \mathbf{y}'||_2^2. \tag{5.5}$$

For M training data observations, the loss function is extended to a cost function that utilizes a mean squared error

$$C = \frac{1}{M}\sum_{i=1}^{M} E(\mathbf{y}_i, \mathbf{y}_i'). \tag{5.6}$$

Change of the loss function with respect to a specific weight is defined by the partial derivative $\frac{\partial E}{\partial w}$ (and $\frac{\partial E}{\partial b}$). Applying the chain rule to the partial derivative is the key to the backpropagation method. For instance, the change in the loss function with respect to weight w_{13}, in Figure 5.1, is

$$\begin{aligned}\frac{\partial E}{\partial w_{13}} &= \frac{\partial E}{\partial y_1} \cdot \frac{\partial y_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_{13}} \\ &= (y_1 - y_1') \cdot 1 \cdot g_1(z_1),\end{aligned} \tag{5.7}$$

where g_3 is assumed to be a unity slope linear activation function. Finding the partial derivative with respect to the weights b_1, w_{11}, w_{21} in the input layer is simply done by further extending (5.7) with the chain rule. In this way, the loss function is referenced to all weights in the ANN, and the cost function is minimized through systematic adjustment of the weights according to an optimization algorithm [*Haykin*, 1999, p. 224]. Variants of the gradient descent method, such as batch gradient descent or stochastic gradient descent, are common choices for minimizing the cost function. This work uses a relatively new optimization algorithm, called a method for stochastic optimization (ADAM) [*Kingma and Ba*, 2015], that has become an extremely popular optimization choice in the ANN community.

5.1.3 High-noon Assumption

Any optimization algorithm for determining the weights of an ANN requires a set of input and output data (training data) to guide the optimization. The Georgia Tech LF Radio Lab

has multiple receivers continuously running, so there is no shortage of magnetic flux density measurements, but the only high-time-resolution measurements of the D region's electron density profile are the highly infrequent and sparse set of sounding rocket measurements. To train the ANN, a large and diverse set of magnetic flux density measurements and the complementary h' and β values must be known.

Performing hundreds of thousands of D region sounding rocket measurements is not practical; it is too costly. Modeling the magnetic flux density at a receiver, given some ionospheric profile, is a feasible task for hundreds of thousands of transmitter to receiver paths and ionospheric electron density profile combinations. This is called synthetic data generation. It is accomplished by first choosing a transmitter to receiver path and a synthetic (fictitious) electron density profile over the path. LWPC is then used to determine B_ϕ at the receiver. This synthesizes the needed training data to train the ANN; $\mathbf{B}_\phi$ is the input to F^{-1}, while $\bar{\mathbf{h}}'$ and $\bar{\boldsymbol{\beta}}$ are the outputs.

But this method is not without challenges, essentially because the synthetic data must match the practically-realizable data, it cannot be idealized in any way, or else the ANN will have been trained with something different than how it is applied. The transmitter power, for example, is not precisely known, as discussed in Chapter 4. Although power levels of certain transmitters do have generally accepted values in the literature, empirically it is clear that the transmitter power varies, likely to the operational needs and capabilities. This presents an obstacle when working with synthetic data. LWPC requires the user to input some transmitter power so that the strength of the field can be calculated. Without knowledge of the radiated power at a given time, the field strength output from LWPC may not correctly align with observations. A similar problem exists with phase, as the transmitter phase is known to not always be steady, as observed in Chapter 3. One possible solution to this is to place a receiver very close to each transmitter, close enough to pick up only the ground wave and not any ionospherically reflected component. Changes in the amplitude or phase of the transmitter is directly observed in the ground wave, and these changes can then be removed from the long distance signal. This is certainly tenuous as it requires a more extensive network of receivers, but even so it is still not foolproof. For example, the phase center of the transmitter, which covers 1000s of acres, may be moving

physically by 10s of meters, which therefore affects the phase differently as a function of direction. The only way to measure and infer that is to have a set of receivers surrounding each transmitter. Furthermore, near field radio wave scattering at the receivers, as discussed in Chapter 4, can impose large biases on measured phase which LWPC is unable to model. These obstacles must be accounted for so that real-world observations can be related to the LWPC model.

McRae and Thomson [2000] present a method to circumvent uncertainties in transmitter phase and power through the use of temporal comparison, where the change in observed and modeled magnetic flux density over some period of time are compared, as opposed to the absolute value. If waveguide parameters $\bar{h}'^0$ and $\bar{\beta}^0$ are known at time t_0, and the observed magnetic flux density is B_ϕ^0, then these measurements can be related to another observation B_ϕ^1 at time t_1. The ratio of the two observations should be equal to the ratio of the modeled field values

$$\frac{B_\phi^1}{B_\phi^0} = \frac{L(\bar{h}'^1, \bar{\beta}^1)}{L(\bar{h}'^0, \bar{\beta}^0)}, \tag{5.8}$$

which is the same as saying the ratio of the amplitudes and the change in phase should be equivalent for the observations and model output, given the correct values for $\bar{h}'^1$ and $\bar{\beta}^1$ are used. Using ratios insulates us from many of these errors, since an unknown multiplier that would scale both the measured values by the same quantity would be removed. Finding these waveguide parameters is done in a similar way to that in (4.4),

$$\left[\bar{h}'^1, \bar{\beta}^1\right] = \underset{h',\beta}{\arg\min} \left\{ \left| \frac{L(h', \beta)}{L(\bar{h}'^0, \bar{\beta}^0)} - \frac{B_\phi^1}{B_\phi^0} \right| \right\}, \tag{5.9}$$

except now the observations are over the same path and the waveguide parameters at times t_0 and t_1 can take on different values.

Equation (5.8) requires knowledge of the B_ϕ^0 observation's waveguide counterparts $\bar{h}'^0$ and $\bar{\beta}^0$. In other words, for each transmitter to receiver path, we need a single reliable connection between the amplitude/phase measurement, and a value of $\bar{h}'$ and $\bar{\beta}$, to 'seed' the rest of the analysis. A good choice for time t_0 is when the Sun is at minimum solar zenith ($t_0 = t_H$), which we refer to as *high-noon*, and the high-noon waveguide parameters

$\bar{h}'^H$ and $\bar{\beta}^H$ are referred to as *seeds*. Empirical measurements of the seeds do not generally exist, but through a minimization technique called seeding (described later in chapter) these values can be estimated. This assumption

$$\frac{B_\phi}{B_\phi^H} = \frac{L(\bar{h}', \bar{\beta})}{L(\bar{h}'^H, \bar{\beta}^H)}, \tag{5.10}$$

is referred to herein as the high-noon assumption, and it is the key to relating observations B_ϕ to the LWPC model $L(\bar{h}', \bar{\beta})$.

Inverse function F^{-1} from (5.2) needs to be modified to allow for the high-noon assumption, giving the new function

$$G^{-1}(\mathbf{B}_\phi, \mathbf{B}_\phi^H) = [\bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}], \tag{5.11}$$

with the extra input vector $\mathbf{B}_\phi^H = \left[B_{\phi 1}^H, \ \ldots \ , B_{\phi N}^H\right]^T$. The ANN is not an exact representation of G^{-1}. Instead, the ANN is an estimate of G^{-1} (referred to as $\widehat{G}^{-1} : \mathbb{C}^N \times \mathbb{C}^N \rightarrow \mathbb{R}^{N\times 2}$), and it is only valid for the range of the training data used to train the ANN. Once the seed values are defined and the weights for the ANN are found, the magnetic flux density observations from a single time step $\mathbf{B}_\phi$ and the high-noon magnetic flux densities from each path $\mathbf{B}_\phi^H$ are input into the estimated inverse function $\widehat{G}^{-1}$, which determines the inferred waveguide parameters $\bar{\mathbf{h}}'$ and $\bar{\boldsymbol{\beta}}$.

A graph of the ANN's topology is shown in Figure 5.2. Notice that the inputs for $\widehat{G}^{-1}$ are defined as complex valued vectors, but each node in the first layer of the ANN only accepts a real value. In most cases, ANNs are constructed with a single real valued input vector and a single real valued output vector, although there are some instances where the vectors may be complex [*Zimmermann et al.*, 2011]. This is why the ANN is designed with a $2N$ real valued input vector (note that ordering of the input and outputs nodes does not matter in this type of topology as long as the ordering is consistent). Inputs to the ANN are the real and imaginary components of the ratio $B_{\phi n}/B_{\phi n}^H$ (this ratio comes from the high-noon assumption), and outputs of the ANN are the average $\bar{h}'_n$ and $\bar{\beta}_n$ values. The figure shown here has a single hidden layer with P nodes, but more layers may be added to

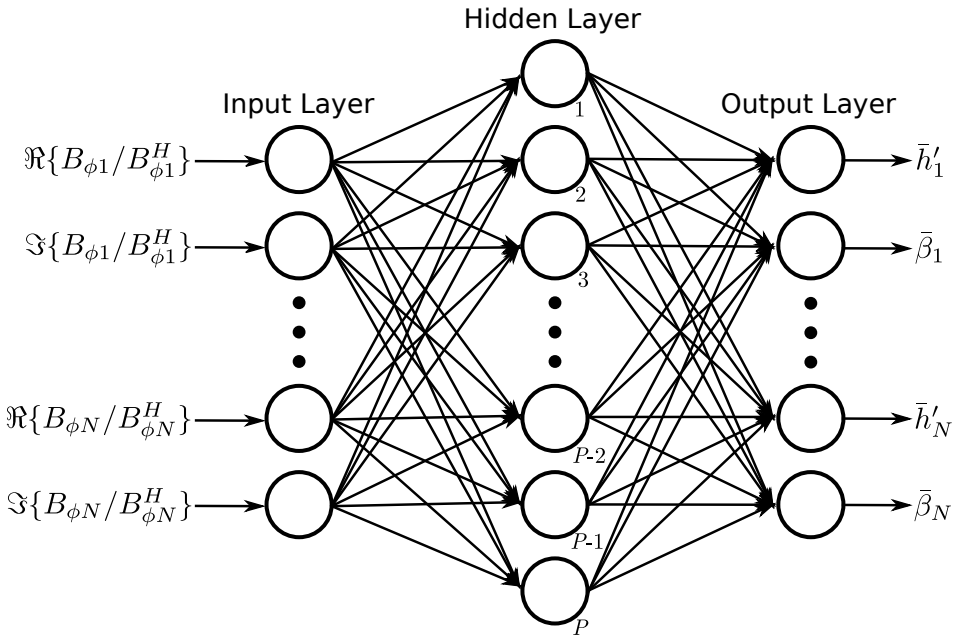

Figure 5.2: Function $\widehat{G}^{-1}$ is an ANN that approximate function G^{-1} in (5.11). Inputs to $\widehat{G}^{-1}$ are observed complex narrowband VLF signals and the complex high-noon coefficients, and the output of the function are the average waveguide parameters.

the network if needed.

There are other advantages to using the high-noon assumption, namely that it intrinsically corrects for imperfect receiver calibration, and also for terrain artifacts that alter the observed amplitude and phase by measurable amounts.

5.2 Synthetic Training Data

The crux for training the ANN is the training data itself. Performance of the ANN will only be as good as the training data used to find the weights of the ANN, so creating a rich and diverse set of synthetic training data is paramount.

5.2.1 Synthetic Data Generation

The purpose of the ANN is to recognize correlations in the waveguide parameters between different paths and to use these correlations to concurrently infer the waveguide parameters for a set of N paths. One approach is to generate a set of random h' and β values for each individual path, but this would produce waveguide parameters that are not correlated

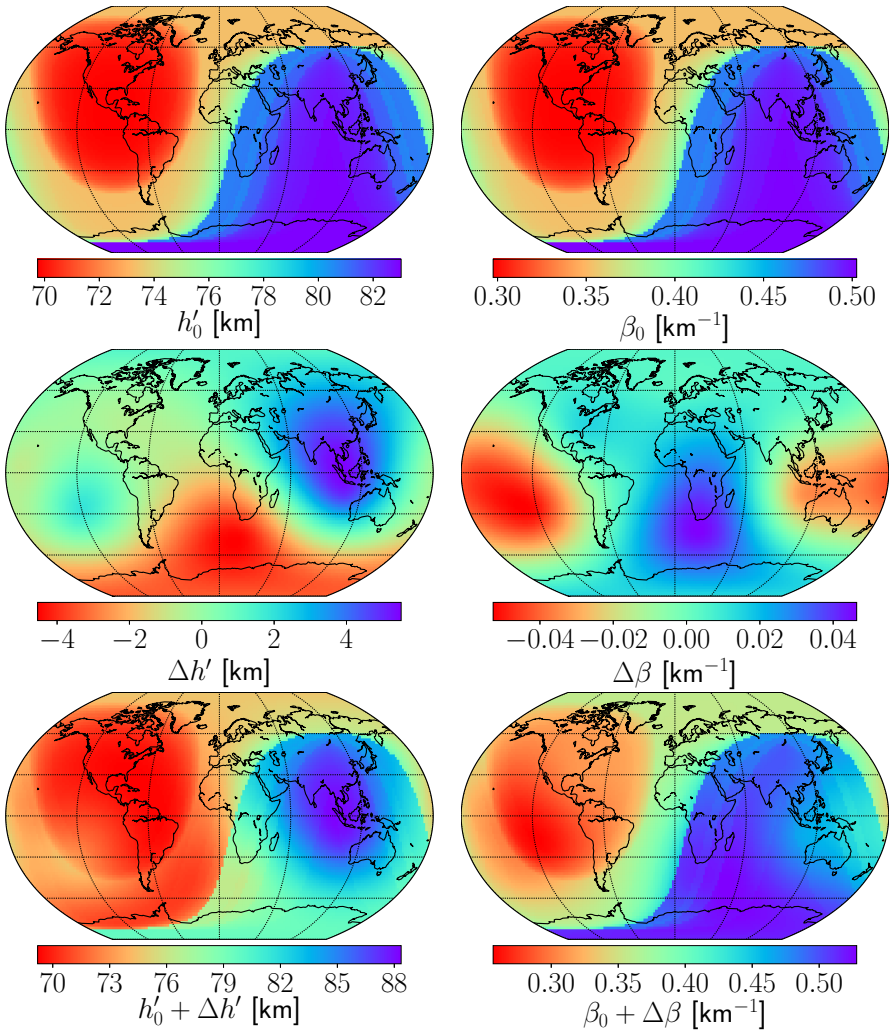

Figure 5.3: Ideal h' and β variations on a global scale. The shape of the deterministic waveguide parameter distribution (top row) comes from IRI on 1 August 2017 at 18:00 UT. Perturbations (middle row) are randomly generated. Deterministic distribution and randomly generated perturbations are combined (bottom row).

between sites. This would defeat the purpose of using the ANN altogether, since we are trying to leverage the fact that different GCPs have at least some correlation in the values between them. Therefore, a single training data step of magnetic flux density measurements and waveguide parameters (the set of $\{\mathbf{B}_\phi, \bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}\}$ for all N paths at a single point in time) must have some correlation between paths. To achieve this, we generate synthetic h' and β values over the entire world. This strategy is referred to here as a *random ionosphere*.

Figure 5.3 shows a random ionosphere example for a single training data step. The random ionosphere is decomposed into two features: the smooth and deterministic ionosphere

(h'_0 and β_0) is shown in the top panels. The middle panels show the random perturbations distributed over the world ($\Delta h'$ and $\Delta\beta$). Together $h'_0 + \Delta h'$ and $\beta_0 + \Delta\beta$ form a random ionosphere for a single time step, shown in the bottom panels. Adding perturbations to the unperturbed ionosphere is necessary so that training data for the ANN is rich with diversity. Having this diversity in the training data teaches the ANN how to respond under non-ideal ionospheric conditions. By extracting the waveguide parameters, for all N paths, from a single random ionosphere, the criteria that waveguide parameters have some dependence between paths is maintained.

The International Reference Ionosphere (IRI) 2007 model [*Bilitza and Reinisch*, 2008] is used to find the deterministic ionosphere in Figure 5.3. There are many different parameters to adjust within IRI, but this work is concerned with using IRI to form a coarse outline for the shape of h'_0 and β_0 over the world. IRI produces an electron density profile, as a function of altitude, for a given latitude, longitude, date, and time. Determination of the waveguide parameters from the electron density profile is not well defined – what altitude range of the profile to consider is subjective. We define the altitude range of interest to be ± 5 km from the effective reflection height (calculated the same way as that in Figure 1.4), and we use a linear least squares fitting technique to determine h' and β. In the h'_0 panel, it is seen that it is daytime in the western hemisphere and nighttime in the eastern hemisphere. Time and date are adjusted within IRI to change the daytime and nighttime locations, along with seasonal variations. Extracting β_0 from IRI does not always produce reasonable results (daytime β_0 values are usually more reasonable than nighttime β_0 values in IRI), so the profile from the h'_0 plot is copied and manually adjusted to give reasonable β_0 values.

The IRI model is known to be generally inaccurate for the D region, due mostly to the lack of available direct measurements. *Friedrich et al.* [2018] augments the IRI model with a sparse set of rocket measurements to improve D region accuracy. What IRI does capture is the rough correlations between different paths that are built into diurnal and seasonal variations. On the other hand, training the ANN with IRI-based synthetic data will bias results from the ANN towards those key features in IRI, which is bad if there is disagreement between the ionosphere's true electron density and these key features in IRI.

An alternative method for random ionosphere generation is to not use IRI and instead use a horizontally homogeneous ionosphere with randomly generated perturbations superimposed on top of it. Waveguide parameters h'_0 and β_0 take on a single random, but physically reasonable, value over the entire world (the top panels in Figure 5.3 would be a constant value/color), and randomly generated perturbations $\Delta h'$ and $\Delta\beta$ are superimposed on top of the homogeneous h'_0 and β_0 values. Essentially, a physically reasonable offset is added to $\Delta h'$ and $\Delta\beta$. This random ionosphere method removes any bias towards assumed diurnal changes, the day/night terminators, the solar cycle, or seasonal variations. This method also removes the ambiguity of inferring h' and β from the IRI electron density profile, and it avoids the problem that IRI sometimes produces unrealistic β values. For this work, we do not use IRI for synthetic random ionosphere generation and instead use this simpler horizontally homogeneous ionosphere with superimposed perturbations method. A diurnal trend analysis of ANN generated waveguide parameters is presented later in this chapter, where it is shown that the ANN infers diurnal trends exceptionally well, even though no diurnal features are included in the synthetic training data.

Creation of the randomized perturbations, Δh and $\Delta\beta$, begins with defining multiple geographic perturbations points and assigning each point a perturbation value that is randomly selected within the range of $-8.0 \leq \Delta h' \leq 8.0$ and $-0.12 \leq \Delta\beta \leq 0.12$. A random number of perturbation points (between 10 and 20) close to evenly distributed around the Earth (a randomly oriented Fibonacci sphere is a good place to start for near even distribution, and then add small random spatial offsets to those points of the Fibonacci sphere) is found to give a good tradeoff between ionospheric diversity and large-scale smoothness. The two homogeneous h_0 and β_0 values (random offsets to $\Delta h'$ and $\Delta\beta$) are randomly selected within the range $68.0 \leq h'_0 \leq 76.0$ and $0.36 \leq \beta_0 \leq 0.44$, and they are then added to Δh and $\Delta\beta$, respectively. Gaps between the points are filled using the spherical interpolation method from *Wessel and Becker* [2008] (the middle panels in Figure 5.3 were produced using this interpolation method).

Ranges for the random values h'_0, β_0, $\Delta h'$, and $\Delta\beta$ are chosen to correspond with those reported in the literature for unperturbed and perturbed daytime ionospheric conditions. The number of perturbation points is limited to the small number of 10 to 20 points because

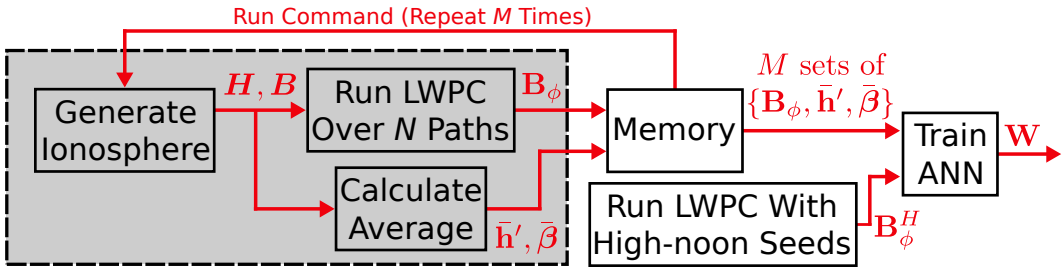

Figure 5.4: Block diagram for synthetic ANN training data generation.

the daytime ionosphere normally does not experience spatially small perturbations. These daytime random value ranges and random number of perturbation points do induce some bias into the synthetic training data, but this bias is less than the bias the IRI model would induce.

There are many parameters to adjust for the random ionosphere generation, and much of it has not been explored. It is suggested that other variations of synthetic data generation be investigated. For example, the random value ranges can also be adjusted for nighttime conditions, where h'_0 and β_0 would need to shift upward and the range of $\Delta\beta$ would need to be expanded. Nighttime conditions would also require an increase in the number of perturbation points to account for the spatially small nighttime perturbations.

Figure 5.4 is a block diagram for the generation of synthetic training data. A random ionosphere is generated to give $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$ vectors for each path for a single training data step ($\boldsymbol{H} = \{\mathbf{h}'_1, \ldots, \mathbf{h}'_N\}$ and $\boldsymbol{B} = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$). The magnetic flux density $B_{\phi n}$, given the associated waveguide parameter vectors $\mathbf{h}'_n$ and $\boldsymbol{\beta}_n$, is then calculated for each path using LWPC. Magnetic flux density calculations and average waveguide parameters are then stored into memory as the set $\{\mathbf{B}_\phi, \bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}\}$. This process is repeated M times ($M = 200,000$ for this work). Once the M^{th} iteration is complete, M random ionosphere sets of $\{\mathbf{B}_\phi, \bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}\}$ are retained in memory (and can be written to file for repeat use). Magnetic flux density values $\mathbf{B}_\phi^H$, for all N paths, are also calculated with LWPC given the chosen seed values ($\bar{h}'^H$ and $\bar{\beta}^H$). With M sets of $\{\mathbf{B}_\phi, \bar{\mathbf{h}}', \bar{\boldsymbol{\beta}}\}$ and $\mathbf{B}_\phi^H$ calculated, the ANN is ready to be trained (i.e., find the set of weights $\mathbf{W}$ that optimize the ANN to approximate the target function G^{-1}). Computing this training data takes about 12 to 24 hours using the Georgia Tech LF Radio Lab's server *Kilaboltz*, which has 64 CPUs and 256 GB of memory.

5.2.2 ANN Performance

There are numerous resources available for training ANNs, and some are better suited for specialized tasks over others, but generally all ANN development packages can handle the simple feed-forward ANN presented here. Most of the code used to generate the synthetic data is written in Python 2.7, so it makes sense to use a python-based ANN package. The Keras API [*Chollet*, 2015] is a useful high-level tool for fast ANN development with the Python language, and it is used in this work for training the ANN. Keras is designed to run on top of a backend engine (low-level optimized tensor manipulation libraries for ANNs), and its purpose is to abstract away from the tediousness of low-level calculations that are of concern only to advance ANN researchers. The backend used in this work is Google's open source TensorFlow library [*Martin et al.*, 2015].

Training an ANN is a slick way of saying that an optimum set of weights $\mathbf{W}$, that minimize the cost function in (5.6), are being determined. Using the Keras API and the synthetic training data, the estimated inverse function $\widehat{G}^{-1}$ is determined by optimizing the weights of the ANN. For this work, it is found that the ANN performs well when it is designed with a single hidden layer of ~300 nodes with ReLU activation function, linear activation functions for the output nodes, and a squared error loss function. Determining how many nodes and layers, and what type of cost and activation functions to use, is known as hyperparameter tuning. There are numerous approaches to hyperparameter tuning [*Bergstra et al.*, 2011], but intuition and thoughtful guessing and checking to minimize the cost function can usually give decent if not good results.

Training and testing of the ANN is done using a 60/20/20 split (60% of the generated data is used for training, 20% for validation, and 20% for testing) of $M = 200,000$ random ionospheres with $N = 18$ transmitter to receiver paths (these paths are shown in Figure 5.6). Training and validation data are used on different neural network topologies for hyperparameter tuning (each ANN topology has a unique set of hyperparameters). All ANN topologies are then tested against a test set that is dedicated for use in the final testing phase (i.e., the test data was not used in any way to train the ANNs). The ANN topology that gives the best testing results is considered to have the optimal hyperparameters for

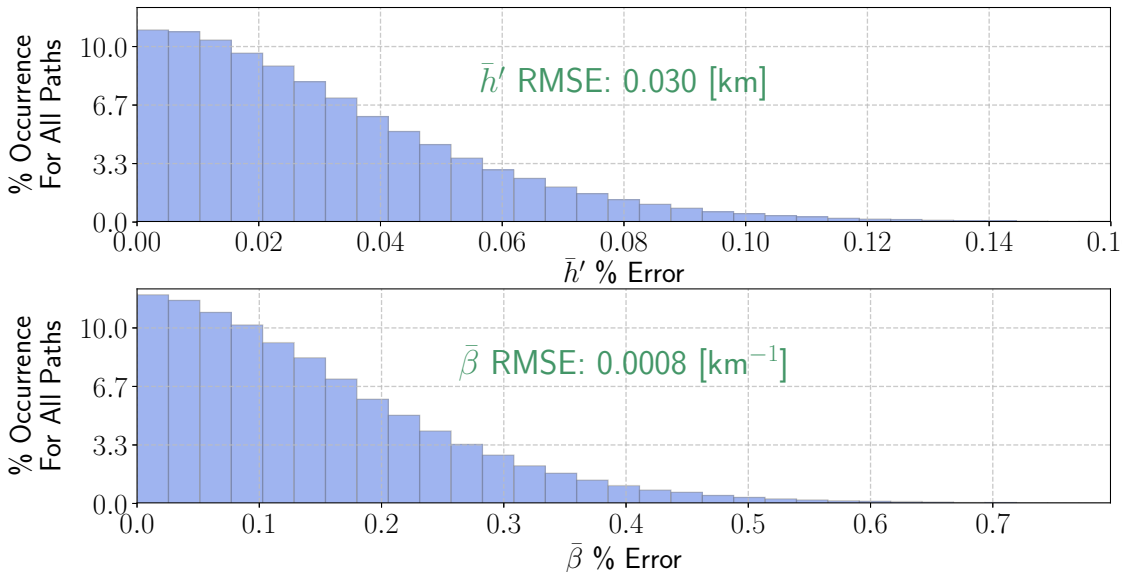

Figure 5.5: ANN performance with synthetic data. The histograms show the aggregate error for 40,000 random ionospheres, each with 18 paths.

estimating inverse function G^{-1}.

Figure 5.5 shows the test data results for the optimal ANN. Seed values of $\bar{h}'^{H} = 71.0$ km and $\beta^{H} = 0.4$ km^{-1} are used to perform this test – these are common values for a quiet daytime ionosphere (see Table 1.1). Only synthetic data is used here, and so similar ANN performance for any reasonable seed values is expected. The two histograms in the figure show the aggregate root-mean-square error (RMSE) for all paths: a mean RMSE of 0.03 km for $\bar{h}'$ and a mean RMSE of 0.0008 km^{-1} for $\bar{\beta}$. Performance of the ANN compares exceptionally well to techniques used in literature. However, this exercise was performed with synthetic data only, and many of the non-ideal characteristics that are measured in observations (e.g., antenna alignment, transmitter source phase uncertainty, changes in transmitted power, and multipath propagation) are not represented within the model. Performance of the ANN is expected to degrade when used with observational data.

5.3 Quiet Day Analysis

Performance testing of the ANN with synthetic data is done merely to tune the ANN and verify the ANN functions properly under ideal conditions. The true performance of the

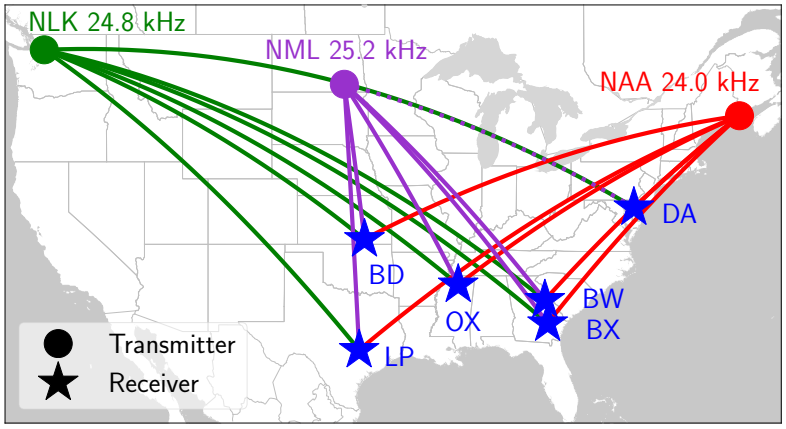

Figure 5.6: Map of narrowband VLF transmitters and Georgia Tech LF Radio Lab receivers. Observations of the three transmitters are made at the six receivers. These observations are used in the ANN to infer the average waveguide parameters along each GCP.

ANN can only be tested with real data. An array of transmitters and receivers are shown in Figure 5.6. This array is a subset of the larger transmitter and receiver array in Figure 1.3 (NAU is removed because of its high-frequency, NPM is removed since it is located far away from the receivers, and PA has strong site errors that make it unsuitable for this work). The ANN framework will be used to infer the average waveguide parameters along $N = 18$ GCPs, formed by the three transmitters and the six receivers.

The daytime ionosphere on 4 October 2017 is found to be extremely quiet, from a solar radiation standpoint, and satellite reports show temperate daytime solar X-ray flux emissions over the continental US. The following analysis uses radio measurements of the three narrowband VLF transmitters starting from 14:00 UT and ending at 22:00 UT, the hours in which all 18 GCPs are 100% daytime. Magnetic flux density $B_{\phi nk}$ is observed every second k for each transmitter to receiver path n. For each second (there are $K = 28,800$ total seconds) all N field measurements are input into the ANN and the inferred $\bar{h}'_{nk}$ and $\bar{\beta}_{nk}$ waveguide parameters, for all paths, are concurrently determined.

5.3.1 Optimal Seed Values

Finding the correct seed values $\bar{h}'^{H}$ and $\bar{\beta}^{H}$ is critical for optimal ANN performance. Seed values are the link that connects real-world observations to synthetic data. When working

with synthetic data only, no connection is required between real-world observations and synthetic data, so it makes sense that any arbitrary seed values (within reason) would give similar results when testing the ANN performance with synthetic data. This is not the case when the ANN is used on real-world observational data. Incorrect seed values will improperly associate the synthetically trained ANN with real-world observations. The method for choosing the optimal seed values with observational data is to train many ANNs with different seed values and see which performs the best on the observational data.

Using the high-noon assumption and equation (5.10), we can calculate the ANN's RMSE over the eight hours of observational data for a given set of seed values,

$$\varepsilon_{\text{seed}}(\bar{h}'^H, \bar{\beta}^H) = \sqrt{\frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K}\left|\frac{B_{\phi nk}}{B^H_{\phi n}} - \frac{L(\bar{h}'_{nk}, \bar{\beta}_{nk})}{L(\bar{h}'^H, \bar{\beta}^H)}\right|^2}. \tag{5.12}$$

Recall that $B^H_{\phi n}$ is the observed magnetic flux density when the Sun is at minimum solar zenith angle at the center point of GCP n, and the $\bar{h}'_{nk}$ and $\bar{\beta}_{nk}$ values come from evaluating $\widehat{G}^{-1}(\mathbf{B}_{\phi k}, \mathbf{B}^H_{\phi}) = [\bar{\mathbf{h}}'_k, \bar{\boldsymbol{\beta}}_k]$.

A minimization method is needed to find the best seed values. The minimization technique used here is to train the ANN for a wide range of reasonable seed values and choose the seed values that give the smallest $\varepsilon_{\text{seed}}$. Figure 5.7 shows $\varepsilon_{\text{seed}}(\bar{h}'^H, \bar{\beta}^H)$ calculated with many different variations of $\bar{h}'^H$ and $\bar{\beta}^H$. A global minimum for $\varepsilon_{\text{seed}}$ is found at $\bar{h}'^H = 70.9$ km and $\bar{\beta}^H = 0.36$ km^{-1}. The seed values at this global minimum are the optimal seed values for the ANN. There is also an interesting low error swath that starts with a low $\bar{h}'^H$ and a high $\bar{\beta}^H$ and then transitions to a middle $\bar{h}'^H$ and a low $\bar{\beta}^H$. The complex value $1/L(\bar{h}'^H, \bar{\beta}^H)$ can be thought of as a tuning coefficient, and adjustment of this coefficient will improve or reduce how well the $\mathbf{B}_{\phi n}$ and $L(\bar{\mathbf{h}}'_n, \bar{\boldsymbol{\beta}}_n)$ temporal curves track each other (this is akin to the temporal comparison method). This swath shows seed values that give good temporal tracking between the $\mathbf{B}_{\phi n}$ and $L(\bar{\mathbf{h}}'_n, \bar{\boldsymbol{\beta}}_n)$ curves, and the swath ultimately converges to the optimal seed values.

Solving for the smallest $\varepsilon_{\text{seed}}$ value is a computationally intensive task. It takes *Kilaboltz* about 48 hours to train the 3,232 ANNs used to make Figure 5.7. For future work, it would be interesting to give each path its own unique set of seed values, allowing for more flexibility

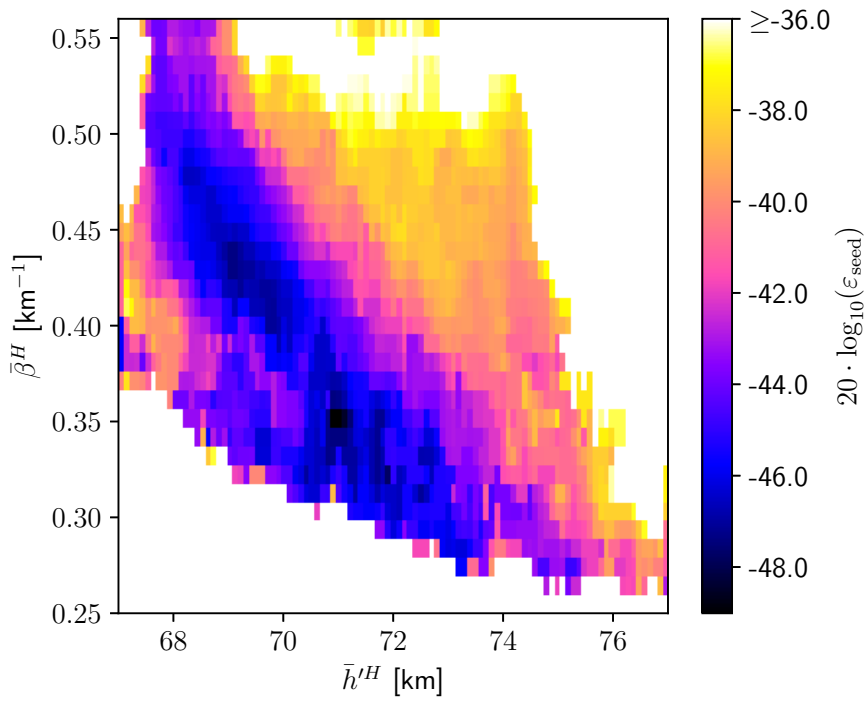

Figure 5.7: Seed error for different seed value combinations. The seed error is calculated using (5.12) and the $N = 18$ observational data sets from 4 October 2017. The optimal (minimum $\varepsilon_{\text{seed}}$) seed values are found to be $\bar{h}'^H = 70.9$ km and $\bar{\beta}^H = 0.36$ km^{-1}.

in the estimate of $\widehat{G}^{-1}$. Having N sets of seed values (one set for each path) would drastically increase the computational burden of finding all N optimal seeding value sets, which is why it is not performed in this work.

Note that the optimal seed values do not define a GCP's $\bar{h}'$ and $\bar{\beta}$ values at high-noon. Even though $B_{\phi n}$ is equivalent to $B^H_{\phi n}$ when path n is at high-noon, all other paths are generally not at high-noon, and $\mathbf{B}_{\phi k} \neq \mathbf{B}^H_{\phi}$ for $k = 1, \ldots, K$ is generally a true statement. This implies that at no single point in time will all observations be equal to their respective high-noon observations, and the ANN will have the flexibility to predict the waveguide parameters for path n that are not necessarily $\bar{h}'^H$ and $\bar{\beta}^H$. If the unlikely event $\mathbf{B}_{\phi k} = \mathbf{B}^H_{\phi}$ does occur, the ANN will predict the waveguide parameters to be the seed values for all paths in the array, but this does not mean the ANN's prediction is wrong.

The optimal seed values are a mechanism to circumvent problems with transmitter amplitude and relative phase, but transmitter amplitude and phase can drift with time. Generally, each day, or even smaller time periods, will have different optimal seed values.

The seed values found here are unique to this day and even to this specific ANN. If the ANN were to be changed through the addition of more nodes, a different number of paths, or even different training data, then the optimal seed values will generally be different.

5.3.2 Inferred Waveguide Parameters

Under daytime conditions with quiet solar activity, the general trends of the waveguide parameters are fairly predictable. As the Sun decreases in solar zenith angle from morning to noon, more solar energy illuminates the overhead D region ionosphere. This causes h' to decrease (a lowering of the effective VLF reflection height) and β to increase (an increase in the gradient of the electron density profile). The opposite occurs when the Sun increases in solar zenith angle from noon to evening; less solar energy is reaching the overhead D region, causing the effective VLF reflection height to increase and sharpness of the electron density profile to decrease.

These general h' and β trends are seen when observational data from 4 October 2017 are used with the $\widehat{G}^{-1}$ function, shown in Figure 5.8. The amplitude and phase observations (shown in the left two columns) are input into the ANN, seeded with $\bar{h}'^{H} = 70.9$ km and $\bar{\beta}^{H} = 0.36$ km^{-1}, and the inferred waveguide parameters $\bar{h}'$ and $\bar{\beta}$ (right two columns) are the output of the ANN. Orange vertical bars show the high-noon time (when the Sun is at minimum solar zenith angle at the midpoint of the path) for each GCP. Each panel has a shaded region which represents the period when the Sun's minimum solar zenith angle was somewhere over the transmitter to receiver path. Most of the amplitude curves have a peak at or near their respective high-noon times, but this is not the case for most of the phase curves, which seem to have more complex curves that do not track well with solar zenith angle. The waveguide parameters all show the general trend that is predicted for daytime conditions with quiet solar activity, and the $\bar{h}'$ nulls and $\bar{\beta}$ peaks are all near their respective high-noon times.

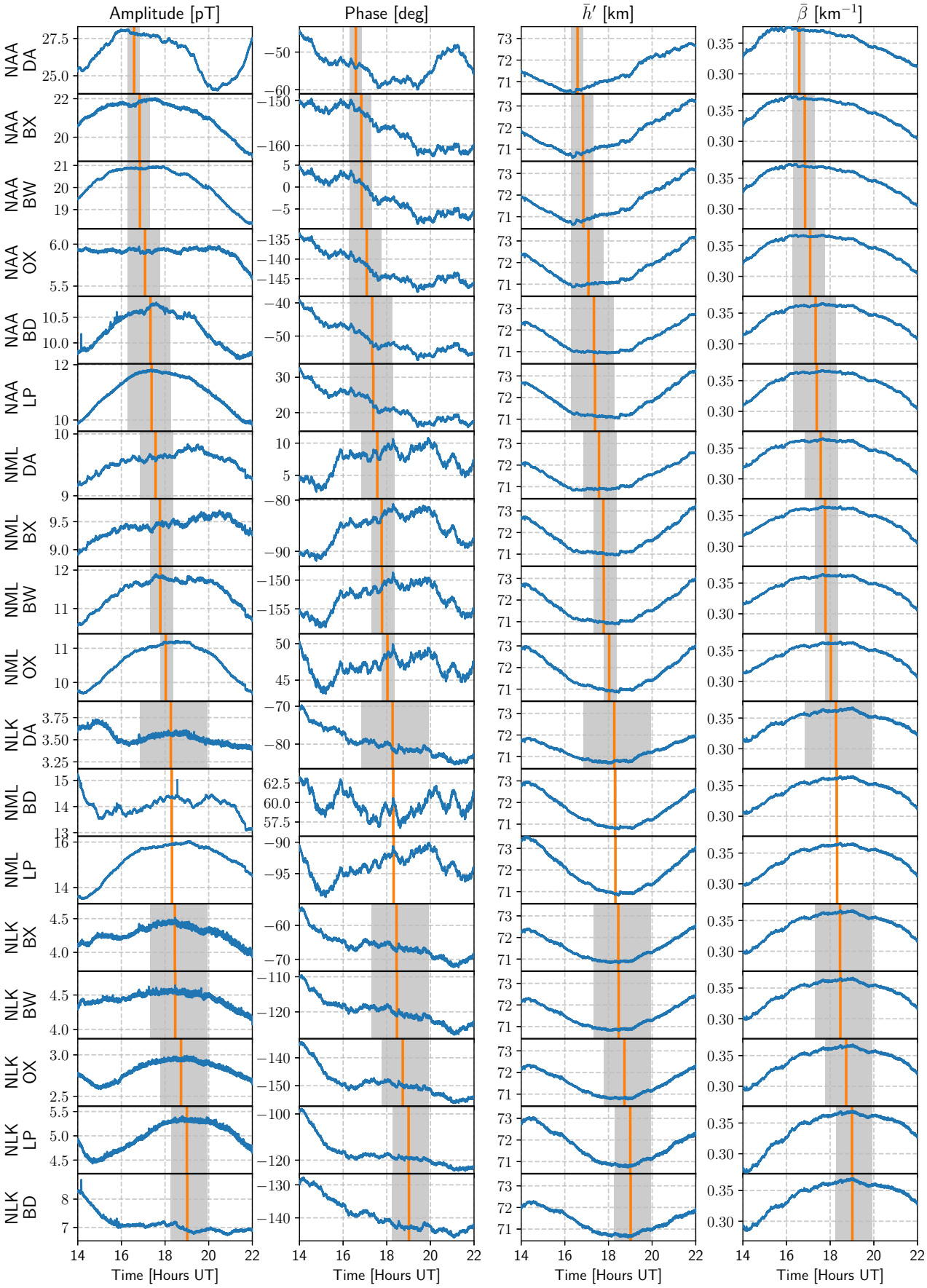

Figure 5.8: Observed amplitude and phase from 4 October 2017 and the inferred $\bar{h}'$ and $\bar{\beta}$ waveguide values found using $\widehat{G}^{-1}$. The vertical orange line is when high-noon occurs, and the shaded region shows how long the Sun was over the path.

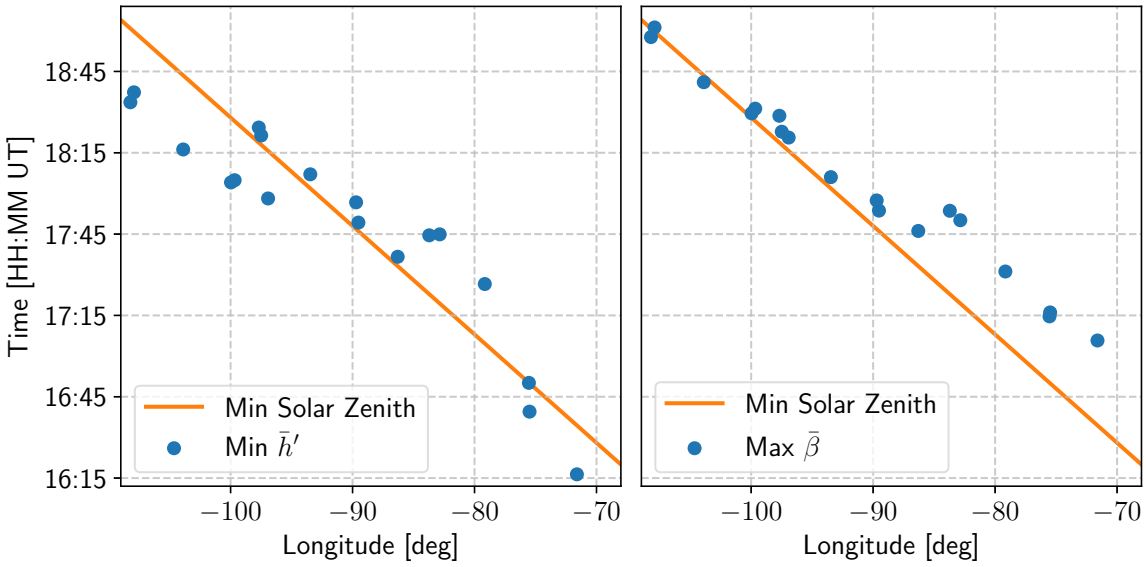

Figure 5.9: Each GCP's longitudinal midpoint is plotted as a blue dot when $\bar{h}'$ is at a minimum and $\bar{\beta}$ is at a maximum in Figure 5.8. The orange line is drawn to show where and when the Sun is at minimum solar zenith angle.

The inferred $\bar{h}'$ and $\bar{\beta}$ curves compare well with previous works. The nulls for $\bar{h}'$ are all near the seed value $\bar{h}'^{H} = 70.9$ km, and the peaks for $\bar{\beta}$ are all near the seed value $\bar{\beta}^{H} = 0.36$ km^{-1}. This is not a coincidence. The inferred waveguide parameters at high-noon will not necessarily be the seeding values, but the chosen seeding values do have a significant impact on the inferred waveguide parameters.

Peaks and nulls of the $\bar{h}'$ and $\bar{\beta}$ curves align well with their respective high-noon times. The ANN itself has no knowledge of the ionosphere's diurnal trends and treats every observation time step independent of all other time steps. This is also true for the synthetic training data. IRI is not used in the generation of synthetic training because it may incorrectly bias how the ANN predicts diurnal trends. Instead, the synthetic training data is generated with a homogeneous random ionosphere with added random perturbations. This means the ANN is not trained to expect any specific type of diurnal trend. Yet, the ANN produces $\bar{h}'$ and $\bar{\beta}$ curves that closely track solar zenith angle.

A second-order polynomial is fit to each of the $\bar{h}'$ and $\bar{\beta}$ curves, and the time associated with the inflection point for each of these parabolic curves is plotted against the longitude of the GCP's midpoint in Figure 5.9. The orange line shows when the Sun is at minimum

solar zenith angle. The ionosphere's chemical response to a change in received solar energy is a non-negligible delay [*Mitra*, 1974], often referred to as *sluggishness* [*Appleton*, 1953]. Thus, the blue dots should lag the orange line just slightly to allow for the ionosphere to reach its maximum ionization level. The blue dots follow the orange line remarkably close, given that the peak/null time is the inflection point of the second-order polynomial fit and the ANN assumes nothing about the temporal variations of daytime ionosphere. Alignment of the second order polynomials' inflection points with the Sun's high-noon position gives confidence that $\widehat{G}^{-1}$ is a good approximation of G^{-1}.

5.3.3 Comparison to Past Works

The general trend of the waveguide parameters in Figure 5.8 appear to be correct, but it is difficult to say how correct they are. One way to verify if the waveguide parameters from $\widehat{G}^{-1}$ are inferred correctly is to compare them to previous works. There are many pitfalls to this approach, and comparison of different works should be done with care. Almost all published works in VLF remote sensing use different geometries of transmitters and receivers. Some GCPs are along dry ground paths, others are almost all sea paths, and many are a mixture of the two. More importantly, the ionosphere is highly dynamic and can exhibit disparate characteristics from one day to the next. These differences are caused by, but not limited to, Earth's seasonal changes and temporal variations in the Sun's behavior. Finally, Earth's background magnetic field, and hence ionosphere, have unique characteristics depending on latitude and longitude. VLF remote sensing literature spans from experiments near the poles down to the equator, and each region has its own unique ionospheric characteristics. Each of these concerns underscore the need for caution when comparing inferred waveguide values between different experimental setups, whether it be later in the day, a different day, a different transmitter and receiver geometry, or different geographic locations.

Figure 5.10 presents inferred waveguide values from many different sources, which include published VLF remote sensing experiments, sounding rocket measurements, and the IRI model. These waveguide parameter curves from different sources are compared against the inferred $\bar{h}'$ and $\bar{\beta}$ curves for the NAA-LP, NML-OX, and NLK-DA GCPs in Figure 5.8.

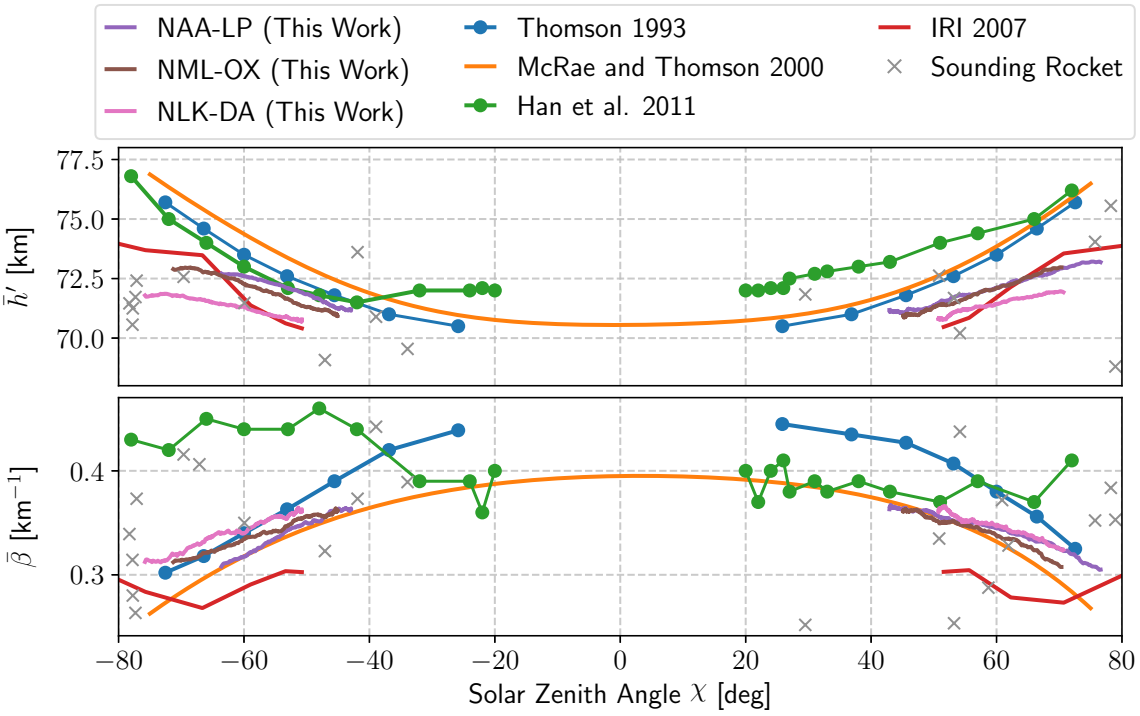

Figure 5.10: Comparison of the inferred waveguide parameters along the NAA-LP, NML-OX, and NLK-DA paths in Figure 5.8 to published experiments, sounding rocket measurements, and the IRI model.

Note that solar zenith angle is the dependent variable here, and a negative solar zenith angle is used for morning, while a positive angle is for afternoon. Solar zenith angle is much simpler to use than time when comparing measurements from different locations or times of the year.

Thomson [1993] inferred $\bar{h}'$ and β using a GCP from the French transmitter FRA to a receiver in Britain and an Australian transmitter NWC to a receiver in Japan. *McRae and Thomson* [2000] used four narrowband VLF transmitters (Omega transmitter in Hawaii, Omega transmitter in Japan, NPM in Hawaii, and NLK in Seattle) and one receiver in Dunedin, New Zealand to infer a single $\bar{h}'$ and β curve for all four GCPs. *Han et al.* [2011] measured sferics emanating from small geographic regions with a receiver in North Carolina during three different days to infer an $\bar{h}'$ and $\bar{\beta}$ curve. $\bar{h}'$ and $\bar{\beta}$ are fit to multiple sounding rocket measurements from *Friedrich et al.* [2018, and references therein] at various solar zenith angles and different locations on Earth. The IRI $\bar{h}'$ and $\bar{\beta}$ curves are determined at

the GCP midpoint location for NML-OX on 4 October.

Overall, the curves in Figure 5.10 have similar trends to those in past works. Some of the sounding rocket measurements do not match with the curves as well, but this mostly occurs at high solar zenith angles where the $\bar{h}'$ and $\bar{\beta}$ assumption begins to break down due to the terminator. The three $\bar{h}'$ curves from this work are lower than the curves from past works, which is surprising because $\bar{h}'$ is typically higher in the winter and lower in the summer. $\bar{h}'$ is lowest along the NLK-DA path and increases slowly as solar zenith angle moves away from zero. This slow increase in $\bar{h}'$ on NLK-DA is expected because the longitudinal distance of the NLK-DA GCP is larger than other paths in Figure 5.6, meaning the Sun is over the path for a longer period of time (the shaded region in Figure 5.8 shows how long the Sun was over each path). This is a drawback to using solar zenith angle, referenced to the center point of the transmitter to receiver path, as an independent variable when comparing different paths – paths with large longitudinal distance will have flatter $\bar{h}'$ and $\bar{\beta}$ curves than those with small longitudinal distance.

$\bar{\beta}$ curves in Figure 5.10 show a little more disagreement than curves for $\bar{h}'$. *Han et al.* [2011] notes that the inference of $\bar{\beta}$ values with sferics produces different results than narrowband VLF remote sensing inferred $\bar{\beta}$ values – namely $\bar{\beta}$ shows little to no daytime solar zenith angle trend when inferring $\bar{\beta}$ with sferics. The $\bar{\beta}$ curves from this work match well with results from *Thomson* [1993] and *McRae and Thomson* [2000] during the morning, but the curve only matches well with *McRae and Thomson* [2000] during the afternoon.

5.3.4 Indirect Error Measurements

Past works have suffered from an inability to quantify the error of waveguide parameter estimation techniques. Measuring the error of these estimation techniques requires waveguide parameter ground truth data, which is almost always unavailable.

Most waveguide parameter estimation experiments contain only one or a couple GCPs. With so few paths, their approach has been to infer the waveguide parameters over a single path, then model the magnetic flux density over the same path using the inferred waveguide parameters, and finally calculate the error between the two fields. Doing so does give an indirect error of the inferred waveguide parameters for this specific path, but this error is

somewhat biased because the inferred waveguide parameters are found by minimizing the modeled and observed error along the same path. A more robust approach is to infer the waveguide parameters along one path and then use the inferred parameters to calculate the indirect error along a second path that was never used to infer the initial waveguide parameters. This second path must be similar enough to the initial path so that both paths can be assumed to have the same ionospheric properties. Reserving a second GCP for error calculations only, and not for initial waveguide parameters inference, is not done in most of the literature due to the limited number of available transmitters and receivers. In this work there are multiple transmitters and receivers, and holding one receiver out of the ANN analysis (so that there are only 15 paths used in $\widehat{G}^{-1}$) will not significantly degrade the inference of the waveguide parameters.

The map of transmitters and receivers in Figure 5.6 show that BW and BX have similar GCPs. During quiet daytime conditions it is valid to approximate the waveguide parameters along the three transmitter to BW paths and the three transmitter to BX paths as equal. A new ANN is formed for $N = 15$ paths, and all three transmitter to BX paths are withheld from the waveguide parameter estimation ($\widehat{G}^{-1} : \mathbb{C}^{15} \times \mathbb{C}^{15} \rightarrow \mathbb{R}^{15\times 2}$). Figure 5.7 shows the optimal seed values for an ANN with $N = 18$, but this new ANN is intended to be formed independent of any BW information. A new set of optimal seed values, that do not use BW paths for calculating $\varepsilon_{\text{seed}}$, are found to be $\bar{h}'^{H} = 70.3$ km and $\bar{\beta}^{H} = 0.33$ km^{-1}. These values are close to the ones found for the $N = 18$ ANN.

Once the ANN is trained, observational data is input into $\widehat{G}^{-1}$ to infer $\bar{h}'$ and $\bar{\beta}$ along the three transmitter to BX GCPs. The inferred BX waveguide parameters are then used in LWPC along the three BW paths and the estimated amplitude and phase at BW is determined for all three GCPs. Figure 5.11 shows the observed and estimated amplitude and phase. Problems arise when directly comparing LWPC's amplitude and phase to observations, due to unknown transmitter power and the lack of relative phase. To avoid this problem, the estimated amplitude and phase, for each GCP, is adjusted so it is equal to the observed amplitude and phase at high-noon. Overall, the predicted amplitudes and phases match very well with the observations. Amplitude RMSEs are below 1 pT and phase RMSEs are all less than 3°. These RMSEs could be reduced even more if some other time than

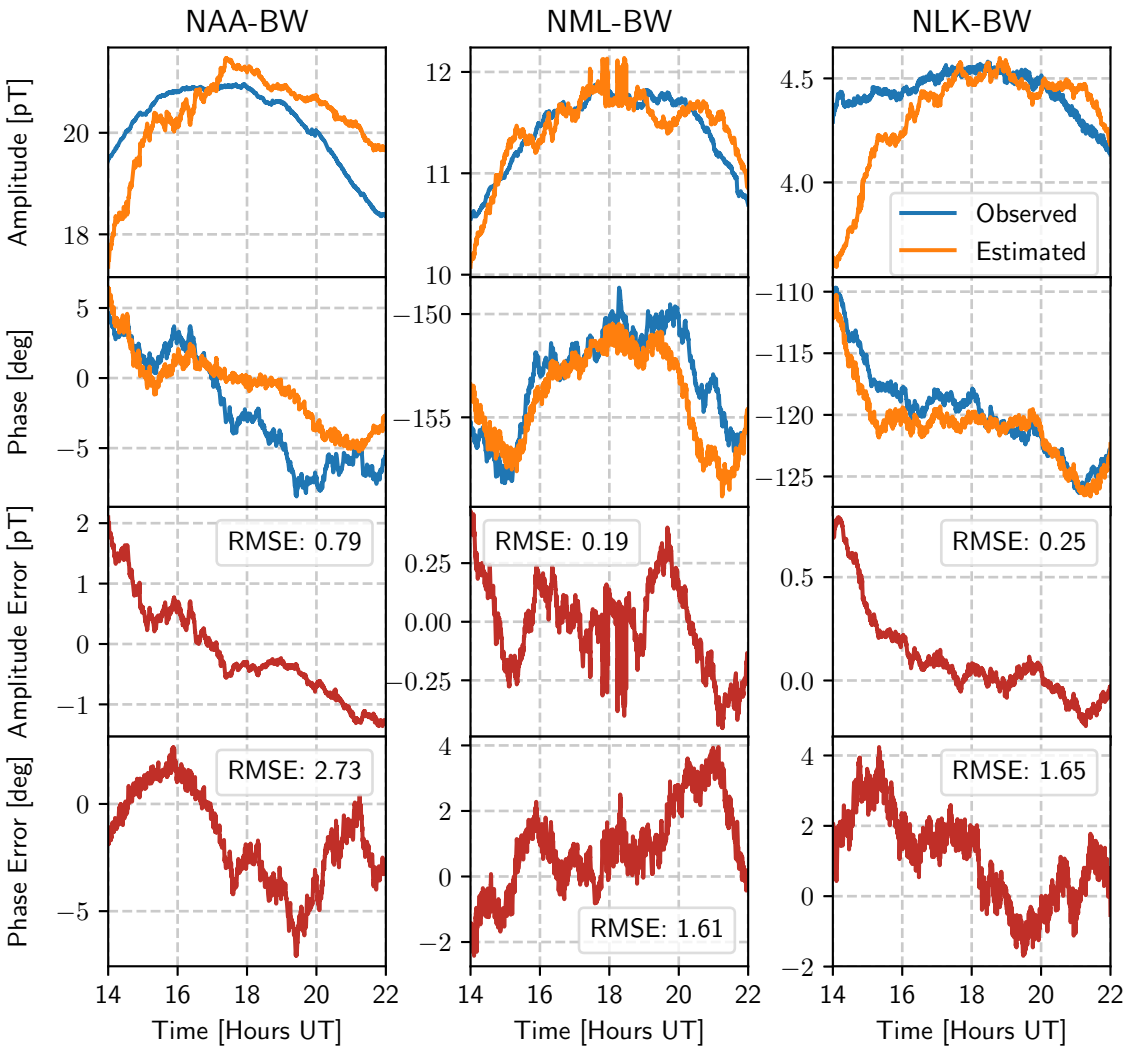

Figure 5.11: Observed and estimated amplitude and phase for the NAA-BW, NML-BW, and NLK-BW paths on 4 October 2017. Estimated amplitudes and phases are found using $\widehat{G}^{-1}$ inferred waveguide parameters from NAA-BX, NML-BX, and NLK-BX paths. Errors between the observed and estimated amplitudes and phase are shown along with the RMSE over the entire 8-hour period.

high-noon is used to adjust the offset of the estimated amplitude and phase but choosing another time would detract from the high-noon assumption used here.

Results for the estimated amplitudes and phases at BW are not too surprising since BX is geographically close to BW. We also want to explore how well we can predict the amplitude and phase at a receiver that is far from other receivers. For this we choose the OX receiver. Looking at the map in Figure 5.6, the closest GCPs for the three transmitters to

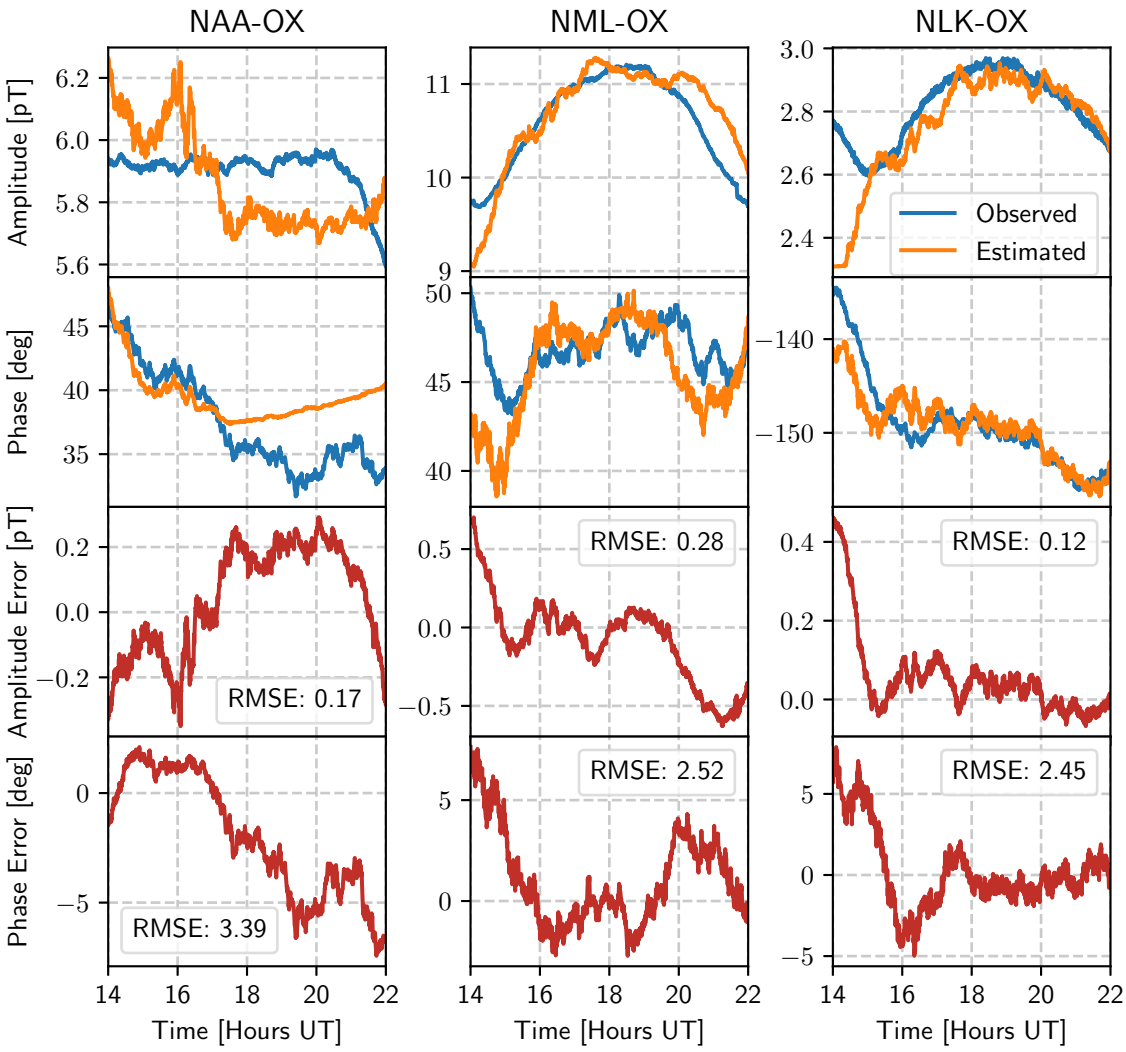

Figure 5.12: Observed and estimated amplitude and phase for the NAA-OX, NML-OX, and NLK-OX paths on 4 October 2017. Estimated amplitudes and phases are found using $\widehat{G}^{-1}$ inferred waveguide parameters from NAA-LP, NML-BX, and NLK-BX paths. Errors between the observed and estimated amplitudes and phase are shown along with the RMSE over the entire 8-hour period.

OX are: NAA-OX is near NAA-LP, NML-OX is near NML-BX, and NLK-OX is near NLK-BX. The LP and BX receivers are quite a far distance from the OX receiver which reduces the confidence that the waveguide parameters are the same between their respective paths. Nevertheless, these pairs of paths will be assumed to have the same waveguide values, and the same approach used to predict the received amplitudes and phases at the BW receiver is again used to predict the received amplitudes and phases at the OX receiver.

Optimal seed values of $\bar{h}'^{H} = 69.8$ km and $\bar{\beta}^{H} = 0.33$ km^{-1} are found for an ANN with the OX receiver removed. The inferred waveguide parameters from $\widehat{G}^{-1}$ along the respective nearby GCPs are used in LWPC to estimate the amplitudes and phases along the three GCPs to OX. Figure 5.12 shows the observed and estimated amplitudes and phases along with the calculated errors. Surprisingly, the amplitude RMSE seems to perform better for the OX predictions than the BW predictions. One reason for this is the OX amplitude error curves have smaller mean offsets than the BW amplitude curves. Overall the amplitude and phase estimates still perform well given that LP and BX are 100s of kilometers away from OX.

5.4 Solar Flare Analysis

The previous section purposefully analyzed a day that had quiet solar activity in an effort to avoid the extra complexity that solar radiation perturbations induce in waveguide parameter inference. Now that it has been shown that $\widehat{G}^{-1}$ performs well in the inference of waveguide parameters during quiet solar day conditions, it is time to explore the effects a solar flare has on the D region and how well $\widehat{G}^{-1}$ handles the extreme change in ionospheric characteristics.

Solar flares produce strong extreme ultraviolet (EUV) radiation, X-ray emissions, and sometimes excess Lyman-α radiation that penetrates the D region ionosphere and causes enhanced ionization [*Tsurutani et al.*, 2009; *Raulin et al.*, 2013]. A C-class solar flare occurred on 15 July 2017 during the daytime over the continental US. X-ray emissions from this flare were measured by the Geostationary Operational Environment Satellite 13 (GOES−13) operated by the National Oceanic and Atmospheric Administration (NOAA) and are shown in the top row of Figure 5.13. GOES−13 has two types of X-ray flux detectors on board, one for short wavelengths (0.5−4 Å) and one for long wavelengths (1−8 Å). Solar X-ray emissions below 8 Å penetrate the furthest into the ionosphere; the shorter the X-ray wavelength, the deeper it will penetrate the ionosphere. Under quiet daytime conditions, solar X-ray flux levels are too low for X-rays to be a dominant D region ionization source. X-ray flux dramatically increases during solar flare events and becomes a dominant driver for enhanced D region ionization [*Mitra*, 1974, p. 60].

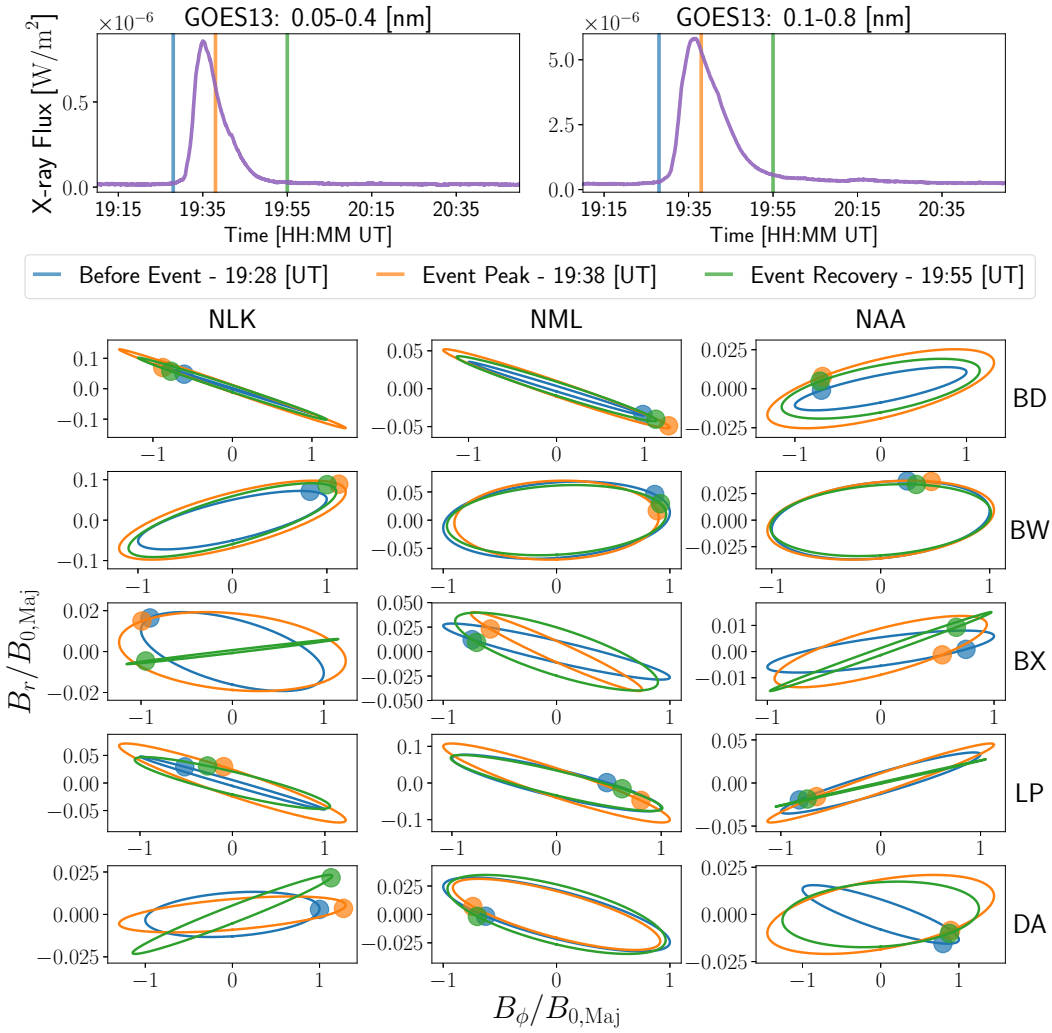

Figure 5.13: GOES−13 X-ray flux measurements (top row) showing a strong C-class solar flare. Ellipse analysis (bottom five rows) is used to show 2-channel narrowband VLF from before (blue), during (orange), and recovery (green) of the solar flare. Ellipses are normalized by the magnetic flux density major axis at 19:28 UT (before event time). Small circles are placed on the ellipses to show start phase.

The X-ray flux curves in Figure 5.13 show that the flare begins at about 19:30 UT and subsides around 19:55 UT. It reaches its peak close to 19:38 UT, which corresponds to the Sun being at minimum solar zenith angle at approximately the geographic longitude line that runs through Salt Lake City, Utah. Recordings of the NLK, NML, and NAA transmitters are made at five different VLF receivers (OX was not operating on this day). The polarization ellipse analysis from Chapter 3 is used to show how the narrowband signals are affected by enhanced ionization of the D region. Narrowband observations just before

the solar flare are shown as blue ellipses, orange ellipses represent narrowband observations when the GOES−13 X-ray flux observation is near maximum, and green ellipses are when the X-ray flux strength is approaching its ambient level prior to the solar flare event. A circle is drawn on each ellipse to represent the start phase, and the magnitude of the ellipses are normalized by the narrowband ellipse major axis values at 19:28 UT ($B_{0,\mathrm{Maj}}$).

Effects of the solar flare widely vary for each transmitter to receiver path. Some of the ellipses change in tilt angle, others change in ellipticity, and some have minimal change. It is clear that the solar flare has a profound impact on the EIWG and the change in observed amplitude and phase is unique to each transmitter and receiver GCP.

5.4.1 Inferred Waveguide Parameters

The same procedure used to determine the 4 October 2017 quiet daytime waveguide parameters in Figure 5.8 is used here for the solar flare event. The seeding method requires the training of many ANNs with different seeding values so that a wide breadth of $\bar{h}'^H$ and $\bar{\beta}^H$ combinations can be considered when searching for the optimal seed values. These ANNs are purely trained with synthetic data and can be recycled for use on different days, provided the transmitter to receiver paths do not change. The OX receiver was unavailable this day, so the ANN used to make Figure 5.8 cannot be used here. However, a whole different set of ANNs, with the OX receiver removed, have already been trained to do the indirect error measurements on the OX receiver in the previous section (Figure 5.12). These ANNs (used to create Figure 5.12) are reused here to calculate the seed errors and give the optimal seed values of $\bar{h}'^H = 70.7$ km and $\bar{\beta}^H = 0.39$ km^{-1}. Conveniently, the last high-noon time is 19:19 UT on the NLK-BD path, which is just before the solar flare. The high-noon assumption would no longer be valid if the event occurred slightly earlier and one or more of the paths had a high-noon time that coincided with the event.

Observed narrowband VLF amplitude and phase and inferred $\bar{h}'$ and $\bar{\beta}$ values are shown in Figure 5.14. The amplitude and phase curves mirror the ellipse analysis in Figure 5.13; the change in amplitude and phase for the solar flare event is unique to each individual GCP. Some amplitudes increase during the flare, some decrease, and some change very little. Even the NAA-BX and NAA-BW paths, which are spatially close to each other, have

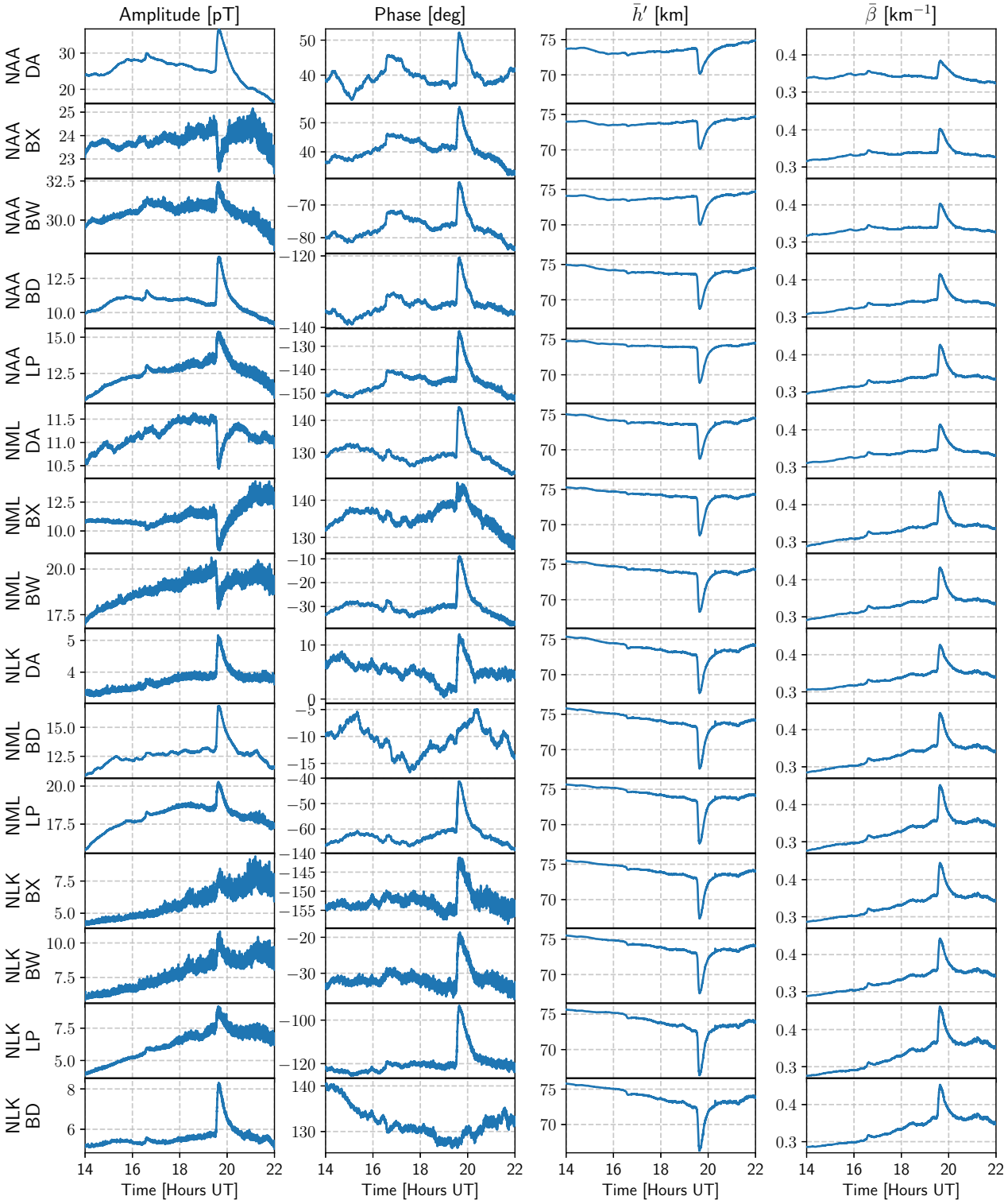

Figure 5.14: Observed amplitude and phase from 15 July 2017 and the inferred $\bar{h}'$ and $\bar{\beta}$ waveguide values found using $\widehat{G}^{-1}$.

different amplitude responses. NAA-BX decreases in amplitude and NAA-BW increases in amplitude. Narrowband phase increases on all paths during the flare, except for NML-BD and NLK-BD, and the amount of phase change is also highly path dependent. The uniqueness to how each narrowband signal responds to the solar flare highlights the complex changes taking place in the EIWG during the event.

Thomson and Clilverd [2001] give a good qualitative discussion on waveguide parameter response to solar flares with excessive X-ray flux levels. Their reasoning is as follows. An increase in D region ionization from increased X-ray flux levels leads to an increase in electron density and lowering of the effective VLF reflection height. This decrease in reflection height is akin to a decrease in h'. During a solar flare, ionization from X-rays typically overshadow ionization levels from Lyman-α and galactic cosmic ray radiation, and the electron density gradient becomes highly dependent on the X-ray energy absorption profile. This ionization profile is generally smooth and sharper than that during quiet daytime conditions. Therefore, β will typically increase during a solar flare. Multiple VLF remote sensing works have shown the general response of the waveguide parameters is for h' to decrease and β to increase in response to a daytime solar flare [e.g., *Thomson and Clilverd*, 2001; *McRae and Thomson*, 2004; *Han and Cummer*, 2010b; *Nina et al.*, 2012; *Singh et al.*, 2014; *Hayes et al.*, 2017].

$\bar{h}'$ and $\bar{\beta}$ curves in Figure 5.14 agree with this qualitative reasoning; all paths show that $\bar{h}'$ decrease and β increases during the event. A median filter is used on the $\bar{h}'$ and $\bar{\beta}$ to clearly show the response of $\widehat{G}^{-1}$ to the event. The amount of change in $\bar{h}'$ and $\bar{\beta}$ is not the same for all paths, but a trend does present itself: change in $\bar{h}'$ and $\bar{\beta}$ tends to be larger for paths that have a GCP midpoint that is further west. This makes sense because the Sun is over the west coast of the US during the solar flare, and paths with the lowest solar zenith angle will generally receive the largest portion of energy from the flare.

5.4.2 Indirect Error Measurements

It is comforting to see in Figure 5.14 that nearby paths with different amplitude and phase curves, such as NAA-BX and NAA-BW, still have similar $\bar{h}'$ and $\bar{\beta}$ curves. The similar change in waveguide parameters for nearby paths comes from the previously stated

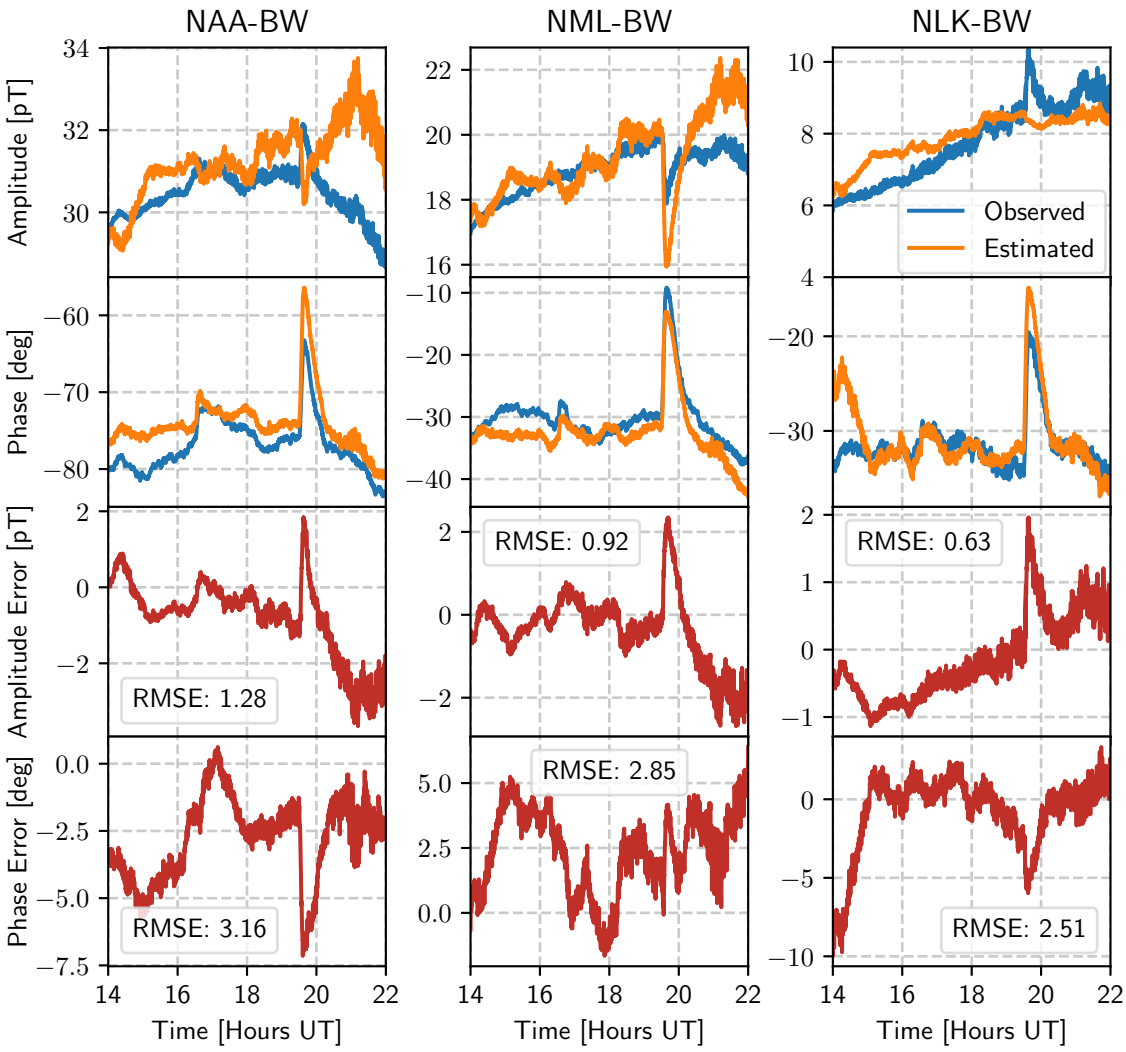

Figure 5.15: Observed and estimated amplitude and phase for the NAA-BW, NML-BW, and NLK-BW paths on 15 July 2017. Estimated amplitudes and phases are found using $\widehat{G}^{-1}$ inferred waveguide parameters from NAA-BX, NML-BX, and NLK-BX paths. Errors between the observed and estimated amplitudes and phase are shown along with the RMSE over the entire 8-hour period.

assumption that the ionosphere will be assumed to have smooth spatial variations. This smooth ionosphere assumption is why large-scale perturbations are used in the generation of synthetic random ionospheres. Thus, the ANN was trained to recognize that nearby paths should have similar waveguide parameters when smooth large-scale perturbations occur. If much smaller perturbations are used in the creation of synthetic data and training of the ANN, then the response of $\widehat{G}^{-1}$ to the solar flare event may be quite different than that

shown here.

An indirect error measurement is made with the BW receiver for the solar flare event, shown in Figure 5.15. This is done in the same manner as the error measurements in Figures 5.11 and 5.12. The OX receiver was not operating on this day, 15 July 2017, and the BW receiver is removed from any analysis involving $\widehat{G}^{-1}$, so the ANN's number of paths is $N = 12$. The seeding method is used to find optimal seed values of $\bar{h}'^{H} = 71.1$ km and $\bar{\beta}^{H} = 0.36$ km^{-1}.

Phase estimates perform remarkably well on all three paths, even during the event. Estimated amplitude on the NML-BW and NLK-BW paths track the observations well. Estimated amplitude for NLK-BW shows almost no response to the event, and estimated NAA-BW amplitude is attenuated during the event even though the observation shows an increase in amplitude. Estimated NML-BW amplitude does show a strong response to the event, in the correct direction, but the amplitude attenuates by about 2 pT more than it should, which is perhaps worse than no response at all.

Three reasons can be given for the poor amplitude RMSE results. First, it may be that the ANN is more focused on phase than amplitude, or at least for solar flare perturbations. This is one of the strongest criticisms of ANNs. ANNs are akin to a black box, and it is near impossible to give a qualitative interpretation as to what the weights at each node mean. Second, the effects of the solar flare on the modal structure of the EIWG are complex. Removing the BW receiver from the ANN has a profound impact with how $\widehat{G}^{-1}$ will respond. Even though the BW and BX receivers are geographically close, it may not be true that $\bar{h}'$ and $\bar{\beta}$ are the same along the two paths under perturbed daytime conditions, which would mean the indirect error method is invalid under perturbed daytime conditions. Third, data from the OX receiver is not available for this day. Having OX data would most certainly improve the ANN's estimates of the waveguide parameters, and the indirect estimates of amplitude and phase at BW would probably improve too.

CHAPTER 6
SUMMARY AND SUGGESTIONS FOR FUTURE WORK

6.1 Summary

VLF remote sensing is one of the primary tools for inferring properties of the D region ionosphere. Yet, D region inference experiments have historically been limited to one or a small number of transmitter to receiver paths. This thesis has expanded upon existing narrowband VLF remote sensing techniques to build the tools needed for a generalized method that concurrently infers the average two-parameter electron density profile along an arbitrary number of transmitter and receiver paths.

In Chapter 1, we introduced VLF radio wave transmission, propagation, and reception. Challenges with the design and construction of narrowband VLF transmitters were examined. We discussed details about the Georgia Tech LF AWESOME receiver and the formation of an array of VLF transmitters and receivers over the continental US. The interesting property that VLF waves propagate to global distances within the Earth-ionosphere waveguide (EIWG) was presented. A review of D region plasma physics and the Appleton-Hartree equation was given. At the end of the chapter we introduce the *Wait and Spies* [1964] two-parameter exponential profile approximation.

In Chapter 2, the theory behind the US Navy's Long-Wavelength Propagation Capability (LWPC) program was introduced. We discussed a mode finding method that considers Earth's ground to be homogeneous and assumes the D region is horizontally homogeneous. Great-circle paths (GCPs) that are inhomogeneous are segmented into smaller homogeneous paths, and mode conversion is used to connect the adjoining segments. This chapter concluded by introducing some VLF remote sensing methods that use waveguide propagation models.

In Chapter 3, we presented some fundamentals for processing narrowband VLF data. We began by detailing the mathematics for minimum-shift keying (MSK) demodulation and the extraction of the narrowband signal's carrier phase and clock phase. A synchronized

demodulation method was introduced to improve signal demodulation on channels that have low SNR. Synchronized demodulation also allows for one-dimensional scattered field analysis [*Dowden et al.*, 1996] to be expanded to a two-dimensional polarization ellipse analysis. Examples of polarization ellipse analysis were given, and it was shown that the ellipse analysis reveals important properties of wave polarization that the one-dimensional scattered field analysis would miss.

In Chapter 4, the EIWG propagation theory in Chapter 2 and narrowband VLF processing techniques in Chapter 3 were joined to define a method for inferring waveguide parameters along a single transmitter to receiver GCP. A method for removing the carrier phase ambiguity between receiver sites was presented, but this method was found to be valid only when the receivers are close to each other. Radio wave scattering from overhead transmission lines was investigated, and it was concluded that nearby scattering objects can induce a bias in the measured carrier phase that exceeds 10°. Radio wave scattering from nearby objects and the carrier phase ambiguity dilemma were found to be problematic when using conventional waveguide parameter inference techniques.

In Chapter 5, a target function was constructed to take in a single time step of narrowband amplitude and phase observations from an arbitrary number of transmitter and receiver combinations and return the inferred waveguide parameters along all paths. The target function was approximated using an artificial neural network (ANN). Synthetic training data was generated using LWPC, which was then used to train the ANN. Real-world performance of the ANN was measured in two ways. First, ANN inferred waveguide parameters were compared to a variety of previously published narrowband VLF remote sensing experiments. Second, ANN inferred waveguide parameters were used in LWPC to predict narrowband VLF amplitude and carrier phase at a receiver that was withheld during waveguide parameter inference. Results showed the approximated target function performs well in capturing temporal and spatial characteristics of the daytime D region.

6.2 Suggestions for Future Work

6.2.1 Polarization Ellipses

Four examples were given for polarization ellipse remote sensing: modal changes due to terminator effects, an Early/fast event, a lightning-induced electron precipitation event, and a solar flare event. Since each of these examples were studied without any statistical information, it is difficult to say if the presented ellipse responses are typical or atypical for each event. A statistical study of each of these event categories, with regard to changes in their polarization ellipse, would be of value to the VLF scientific community.

6.2.2 Polarization Ellipsoids

Under normal circumstances, the magnetic flux density of a VLF wave propagating in the EIWG is strongest in the horizontal plane, but still, the vertical magnetic flux density is not always negligible [*Silber et al.*, 2015]. Working with this vertical field is often quite difficult because of its low SNR. It was shown that the synchronized narrowband demodulation technique is useful when one channel has a low SNR, and this technique can be extended to the vertical magnetic flux density to improve its carrier phase estimate. Addition of the vertical component would extend the polarization ellipse technique to a polarization ellipsoid technique. The utility of polarization ellipsoid analysis would further extend our understanding of VLF EIWG propagation.

6.2.3 Carrier Phase Ambiguity

The easiest way, from a processing standpoint, to remove the carrier phase ambiguity is to simply decrease the transmitter's frequency (make the wavelength longer). This, of course, is not possible. The analogous approach to increasing the wavelength would be to reduce the distance between receivers. This could be done by moving the currently deployed LF AWESOME receivers closer together or by filling in the gaps in the receiver array with new receivers.

6.2.4 Radio Wave Scattering

Unfortunately, radio wave scattering from nearby objects is a challenge that the VLF radio community will always face. Little information about VLF radio wave scattering is available in the literature, and a deeper investigation is needed to understand the impact this scattering effect has on VLF measurements.

6.2.5 Improvements to the ANN Method

In Chapter 5, an ANN was used to approximate the target function G^{-1}. The ANN approximation method has an enormous amount of flexibility and further experimentation with this method is suggested. The following are some suggested ANN changes to explore:

- This thesis has limited all waveguide parameter inference to be the average waveguide parameters along the transmitter to receiver path. This assumption is typically valid for total daytime paths, but it breaks down when the terminator is over the path or during nighttime. If the ANN method is to be extended to morning, evening, or nighttime periods, then some polynomial or possibly piece-wise fit for the waveguide parameters is needed. Extending beyond the average waveguide parameter assumption will increase the complexity of the ANN, and it will possibly require more transmitter to receiver paths since more unknown variables are being solved for.

- Performance of the ANN is limited to the quality of the training data. There may be some value to using IRI [*Bilitza and Reinisch*, 2008] or FIRI [*Friedrich et al.*, 2018] for synthetic data generation. Doing so is tricky because it is difficult to tell if adding these deterministic biases in the synthetic data are improving or diminishing the performance of the ANN. It is also worth looking into adjustment of the shape, distribution, and intensity of the synthetic data perturbations ($\Delta h'$ and $\Delta\beta$).

- The high-noon assumption was needed to connect the synthetically trained ANN to real-world observations. This assumption restricts the seed values to be the same for all paths. There may be some performance improvement with the ANN if the seed values were given the freedom to take on different values along each path. Doing so,

however, will greatly increase the computational burden for finding the optimized seed values.

- Analysis of nighttime observations is a logical next step for the work presented in this thesis. A few considerations must be made to account for nighttime conditions. First, the synthetic training data will need to be adjusted. This most likely means more small scale perturbations need to be added to the random ionosphere generation, which in turn may require additional neurons and layers in the ANN. Second, radio wave scattering from small scale ionospheric perturbations, like Early/fast events, occur during the nighttime. LWPC models propagation along a two-dimensional path and is unable to account for three-dimensional radio wave scattering. Finally, it does not make sense to use the high-noon assumption during the nighttime, since the sun never reaches its minimum solar zenith angle during this period. It may be possible to use the high-noon seed values from the previous and/or next day. If not, then a new method for relating synthetic and observational data must be developed.

- This work has presented a method for combining current VLF observations to improve the inference of ionospheric properties. But the type of data used does not have to be limited to VLF observations. Other data types, such as total electron content (TEC), electron density profiles from ionosondes, sounding rocket electron density profiles, solar X-ray flux, and sunspot number, are also useful when inferring ionospheric properties. *Scherliess et al.* [2017] uses an ensemble Kalman filter to combine physical models and numerous data observation types for ionospheric parameter estimation in their Global Assimilation on Ionospheric Measurements Full Physics (GAIM-FP) model. A similar ensemble Kalman filter method could be used with the ANN method presented here along with other types of observations to improve the inference of ionospheric parameters. An added advantage to the ensemble Kalman filter method is it adds a temporal component to the inference of ionospheric parameters, something the ANN method does not currently do.

- The azimuthal and radial components of the magnetic flux density are both useful measurements for VLF remote sensing, as seen in the polarization ellipse analysis. It

is suggested that the azimuthal and radial components both be used within the ANN.

- An array of three transmitters and six receivers were used in Chapter 5. Adding more transmitters and receivers to the array would most certainly improve the performance of the ANN. The NPM (Hawaii) and NAU (Puerto Rico) transmitters are the two closest VLF transmitters to the continental US. Adding NPM would greatly reduce the period for when it is daytime over all GCPs in the array. NAU operates at a frequency much higher (40.8 kHz) than NAA, NML, and NLK, which is not necessarily a bad thing, but the impact of using such a different frequency is unknown. Adding receivers around the OX region and closer to the DA would give more uniform GCP coverage which would likely improve performance of the ANN, and this denser set of receivers would allow for more combinations of indirect error measurements.

BIBLIOGRAPHY

Aikin, A. C., J. A. Kane, and J. Troim (1964), Some results of rocket experiments in the quiet D region, *Journal of Geophysical Research*, *69*(21), 4621–4628, doi:10.1029/JZ069i021p04621.

Appleton, E. V. (1953), A note on the sluggishness of the ionosphere, *Journal of Atmospheric and Terrestrial Physics*, *3*(5), 282–284, doi:10.1016/0021-9169(53)90129-9.

Bainbridge, G., and U. S. Inan (2003), Ionospheric D region electron density profiles derived from the measured interference pattern of VLF waveguide modes, *Radio Science*, *38*(4), doi:10.1029/2002RS002686.

Barr, R., D. L. Jones, and C. J. Rodger (2000), ELF and VLF radio waves, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*, 1689–1718.

Barron, D. W., and K. G. Budden (1959), The Numerical Solution of Differential Equations Governing the Reflexion of Long Radio Waves from the Ionosphere. III, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *249*(1258), 387–401, doi:10.1098/rspa.1959.0031.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl (2011), Algorithms for hyper-parameter optimization, *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2546–2554.

Bickel, J. E., J. A. Ferguson, and G. V. Stanley (1970), Experimental Observation of Magnetic Field Effects on VLF Propagation at Night, *Radio Science*, *5*(1), 19–25, doi:10.1029/RS005i001p00019.

Bilitza, D., and B. Reinisch (2008), International Reference Ionosphere 2007: Improvements and new parameters, *Advances in Space Research*, *42*(4), 599–609, doi:10.1016/J.ASR.2007.07.048.

Bittencourt, J. A. (2004), *Fundamentals of Plasma Physics*, Springer New York, New York, NY, doi:10.1007/978-1-4757-4030-1.

Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman (2001), Combined Satellite and Surface-Based Estimation of the Intracloud-Cloud-to-Ground Lightning Ratio over the Continental United States, *Monthly Weather Review*, *129*(1), 108–122, doi:10.1175/1520-0493(2001)129⟨0108:CSASBE⟩2.0.CO;2.

Booker, H. G. (1936), Oblique Propagation of Electromagnetic Waves in a Slowly-Varying Non-Isotropic Medium, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *155*(885), 235–257, doi:10.1098/rspa.1936.0096.

Budden, K. G. (1955a), The Numerical Solution of Differential Equations Governing Reflexion of Long Radio Waves from the Ionosphere, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *227*(1171), 516–537, doi:10.1098/rspa.1955.0027.

Budden, K. G. (1955b), The Numerical Solution of the Differential Equations Governing the Reflexion of Long Radio Waves from the Ionosphere. II, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *248*(939), 45–72, doi:10.1098/rsta.1955.0009.

Budden, K. G. (1961), *The wave-guide mode theory of wave propagation*, 325 pp., Logos Press, London.

Budden, K. G. (1962), The influence of the earth's magnetic field on radio propagation by wave-guide modes, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *265*(1323), 538–553, doi:10.1098/rspa.1962.0041.

Budden, K. G. (1985), *The Propagation of Radio Waves: The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere*, Cambridge University Press, doi: 10.1017/CBO9780511564321.

Budden, K. G., and H. G. Martin (1962), The Ionosphere as a Whispering Gallery, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *265*(1323), 554–569, doi:10.1098/rspa.1962.0042.

Carlson, A. B., P. B. Crilly, and J. C. Rutledge (2002), *Communication systems: an introduction to signals and noise in electrical communication*, 4th ed., 850 pp., McGraw-Hill, Boston.

Carpenter, D., T. Bell, and A. Smith (1988), The Siple VLF transmitter as a multi-frequency probe of the Earth-ionosphere waveguide, *Journal of Atmospheric and Terrestrial Physics*, *50*(2), 105–115, doi:10.1016/0021-9169(88)90048-7.

CCIR (1990), Radio Propagation and circuit performance at frequencies below about 30 kHz, *Tech. rep.*, Committee International Radio Consultative, Dusseldorf.

Chen, F. F. (1984), *Introduction to plasma physics and controlled fusion*, Plenum Press.

Chevalier, M. W., and U. S. Inan (2006), A Technique for Efficiently Modeling Long-Path Propagation for Use in Both FDFD and FDTD, *IEEE Antennas and Wireless Propagation Letters*, *5*, 525–528, doi:10.1109/LAWP.2006.887551.

Chollet, F. (2015), Keras, https://github.com/fchollet/keras.

Chrissan, D. A., and A. C. Fraser-Smith (1996), Seasonal variations of globally measured ELF/VLF radio noise, *Radio Science*, *31*(5), 1141–1152, doi:10.1029/96RS01930.

Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, and M. F. Stewart (2003), Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *Journal of Geophysical Research*, *108*(D1), 4005, doi:10.1029/2002JD002347.

Chulliat, A., S. Macmillan, P. Alken, C. Beggan, M. Nair, B. Hamilton, A. Woods, V. Ridley, S. Maus, and A. Thomson (2014), The US/UK World Magnetic Model for 2015-2020, *Tech. rep.*, NOAA National Geophysical Data Center, Boulder, CO, doi: 10.7289/V5TH8JNW.

Clemmow, P. C., and J. Heading (1954), Coupled forms of the differential equations governing radio propagation in the ionosphere, *Mathematical Proceedings of the Cambridge Philosophical Society*, *50*(02), 319, doi:10.1017/S030500410002939X.

Clilverd, M. A., N. R. Thomson, and C. J. Rodger (1999), Sunrise effects on VLF signals propagating over a long north-south path, *Radio Science*, *34*(4), 939–948, doi:10.1029/1999RS900052.

Cohen, M., U. Inan, and E. Paschal (2010a), Sensitive Broadband ELF/VLF Radio Reception With the AWESOME Instrument, *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 3–17, doi:10.1109/TGRS.2009.2028334.

Cohen, M. B., R. K. Said, and U. S. Inan (2010b), Mitigation of 50-60 Hz power line interference in geophysical data, *Radio Science*, *45*(6), doi:10.1029/2010RS004420.

Cohen, M. B., R. K. Said, E. W. Paschal, J. C. McCormick, N. C. Gross, L. Thompson, M. Higginson-Rollins, U. S. Inan, and J. Chang (2018a), Broadband longwave radio remote sensing instrumentation, *Review of Scientific Instruments*, *89*(9), doi:10.1063/1.5041419.

Cohen, M. B., N. C. Gross, R. A. Marshall, M. Golkowski, W. Liles, D. Rodriguez, and J. Rockway (2018b), Lower Ionospheric Response to the 2017 Great American Solar Eclipse, *Geophysical Research Letters*, *45*(8), 3348–3355.

Cotts, B. R. T. (2011), Global quantification of lightning-induced electron precipitation using very low frequency remote sensing, Ph.D. thesis, Stanford University.

Cotts, B. R. T., and U. S. Inan (2007), VLF observation of long ionospheric recovery events, *Geophysical Research Letters*, *34*(14), doi:10.1029/2007GL030094.

Crombie, D. D. (1961), Reflection From a Sharply Bounded Ionosphere for VLF Propagation Perpendicular to the Magnetic Meridian, *Journal of Research of the National Bureau of Standards*, *65*(5).

Crombie, D. D. (1964), Periodic Fading of VLF Signals Received Over Long Paths During Sunrise and Sunset, *Radio Science Journal of Research*, *680*(I), 27–34.

Cummer, S. A. (1997), Lightning and ionospheric remote sensing using VLF/LF radio atmospherics, Ph.D. thesis, Stanford University.

Cummer, S. A., U. S. Inan, and T. F. Bell (1998), Ionospheric D region remote sensing using VLF radio atmospherics, *Radio Science*, *33*(6), 1781–1792, doi:10.1029/98RS02381.

Cummins, K. L., and M. J. Murphy (2009), An Overview of Lightning Locating Systems: History, Techniques, and Data Uses, With an In-Depth Look at the U.S. NLDN, *IEEE Transactions on Electromagnetic Compatibility*, *51*(3), 499–518, doi:10.1109/TEMC.2009.2023450.

de Lisle, J. M. (1967), A Note on the Reflection Coefficient of a Sharply Bounded Ionosphere for VLF Signals at the Magnetic Equator, *Radio Science*, *2*(6), 653–657, doi:10.1002/rds196726653.

Dowden, R. L., J. B. Brundell, and W. A. Lyons (1996), Are VLF rapid onset, rapid decay perturbations produced by scattering off sprite plasma?, *Journal of Geophysical Research: Atmospheres*, *101*(D14), 19,175–19,183, doi:10.1029/96JD01346.

Dowden, R. L., J. B. Brundell, and C. J. Rodger (1997), Temporal evolution of very strong Trimpis observed at Darwin, Australia, *Geophysical Research Letters*, *24*(19), 2419–2422, doi:10.1029/97GL02357.

Ferguson, J. (1992), A Review of the Ionospheric Model for the Long Wave Prediction Capability, *Tech. rep.*, Naval Command, Control and Ocean Surveillance Center, Sand Diego, CA.

Ferguson, J. A. (1980), Ionospheric profiles for prediciting nighttime VLF/LF propagation, *Tech. rep.*, Naval Ocean Systems Center, Springfield, VA.

Ferguson, J. A. (1998), Computer programs for assessment of long- wavelength radio communications, version 2.0, *Tech. rep.*, Space and Naval Warfare Systems Center, San Diego, CA.

Ferguson, J. A., and F. P. Snyder (1980), Approximate vlf/lf waveguide mode conversion Model. Computer Applications: FASTMC and BUMP, *Tech. rep.*, Naval Ocean Systems Center, San Diego.

Francey, R. J. (1970), Electron production in the ionospheric D region by cosmic X rays, *Journal of Geophysical Research*, *75*(25), 4849–4862, doi:10.1029/JA075i025p04849.

Friedrich, M., C. Pock, and K. Torkar (2018), FIRI-2018, an Updated Empirical Model of the Lower Ionosphere, *Journal of Geophysical Research: Space Physics*, doi:10.1029/2018JA025437.

Füllekrug, M., A. Mezentsev, R. Watson, S. Gaffet, I. Astin, and A. Evans (2014), Array analysis of electromagnetic radiation from radio transmitters for submarine communication, *Geophysical Research Letters*, *41*(24), 9143–9149, doi:10.1002/2014GL062126.

Galejs, J. (1971), VLF propagation across discontinuous daytime-to nighttime transitions in anisotropic terrestrial waveguide, *IEEE Transactions on Antennas and Propagation*, *19*(6), 756–762, doi:10.1109/TAP.1971.1140034.

Galejs, J. (1972), *Terrestrial propagation of long electromagnetic waves*, 362 pp., Pergamon Press.

Gobbel, J. T. (1967), *An aid for the optimum design of antenna systems for very low frequencies*, Monterey, California Naval Postgraduate School.

Gołkowski, M., and U. S. Inan (2008), Multistation observations of ELF/VLF whistler mode chorus, *Journal of Geophysical Research: Space Physics*, *113*(A8), doi:10.1029/2007JA012977.

Gołkowski, M., N. C. Gross, R. C. Moore, B. R. T. Cotts, and M. Mitchell (2014), Observation of local and conjugate ionospheric perturbations from individual oceanic lightning flashes, *Geophysical Research Letters*, *41*(2), 273–279, doi:10.1002/2013GL058861.

Gross, N. C., M. B. Cohen, R. K. Said, and M. Gołkowski (2018), Polarization of Narrowband VLF Transmitter Signals as an Ionospheric Diagnostic, *Journal of Geophysical Research: Space Physics*, *123*(1), 901–917, doi:10.1002/2017JA024907.

Haldoupis, C., M. Cohen, E. Arnone, B. Cotts, and S. Dietrich (2013), The VLF fingerprint of elves: Step-like and long-recovery early VLF perturbations caused by powerful CG lightning EM pulses, *Journal of Geophysical Research: Space Physics*, *118*(8), 5392–5402, doi:10.1002/jgra.50489.

Han, F., and S. A. Cummer (2010a), Midlatitude nighttime D region ionosphere variability on hourly to monthly time scales, *Journal of Geophysical Research: Space Physics*, *115*(A9), doi:10.1029/2010JA015437.

Han, F., and S. A. Cummer (2010b), Midlatitude daytime D region ionosphere variations measured from radio atmospherics, *Journal of Geophysical Research: Space Physics*, *115*(A10), doi:10.1029/2010JA015715.

Han, F., S. A. Cummer, J. Li, and G. Lu (2011), Daytime ionospheric D region sharpness derived from VLF radio atmospherics, *Journal of Geophysical Research: Space Physics*, *116*(A5), doi:10.1029/2010JA016299.

Harrington, P. (2012), *Machine learning in action*, 354 pp., Manning Publications.

Hastie, T., R. Tibshirani, and J. H. Friedman (2001), *The elements of statistical learning : data mining, inference, and prediction*, 533 pp., Springer.

Hayes, L. A., P. T. Gallagher, J. McCauley, B. R. Dennis, J. Ireland, and A. Inglis (2017), Pulsations in the Earth's Lower Ionosphere Synchronized With Solar Flare Emission, *Journal of Geophysical Research: Space Physics*, *122*(10), 9841–9847, doi:10.1002/2017JA024647.

Haykin, S. S. (1999), *Neural networks: a comprehensive foundation*, 2nd ed., 842 pp., Prentice Hall.

Hiscox, W. L., E. P. Krider, A. E. Pifer, and M. A. Uman (1984), A systematic method for identifying and correcting "site errors" in a network of magnetic direction finders, in *International Aerospace and Ground Conference on Lightning and Static Electricity*, Orlando, FL.

Holle, R. L., K. L. Cummins, and W. A. Brooks (2016), Seasonal, Monthly, and Weekly Distributions of NLDN and GLD360 Cloud-to-Ground Lightning, *Monthly Weather Review*, *144*(8), 2855–2870, doi:10.1175/MWR-D-16-0051.1.

Horner, F. (1954), The accuracy of the location of sources of atmospherics by radio direction-finding, *Proceedings of the IEE - Part III: Radio and Communication Engineering*, *101*(74), 383–390, doi:10.1049/pi-3.1954.0091.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*(5), 359–366, doi:10.1016/0893-6080(89)90020-8.

Hosseini, P., M. Gołkowski, and D. L. Turner (2017), Unique concurrent observations of whistler mode hiss, chorus, and triggered emissions, *Journal of Geophysical Research: Space Physics*, *122*(6), 6271–6282, doi:10.1002/2017JA024072.

Hu, W., and S. Cummer (2006), An FDTD Model for Low and High Altitude Lightning-Generated EM Fields, *IEEE Transactions on Antennas and Propagation*, *54*(5), 1513–1522, doi:10.1109/TAP.2006.874336.

Inan, U. S., and M. Golkowski (2011), *Principles of plasma physics for engineers and scientists*, 270 pp., Cambridge University Press.

Inan, U. S., and A. S. Inan (2000), *Electromagnetic Waves*, 800 pp., Prentice Hall.

Inan, U. S., V. P. Pasko, and T. F. Bell (1996), Sustained heating of the ionosphere above thunderstorms as evidenced in early/fast VLF events, *Geophysical Research Letters*, *23*(10), 1067–1070, doi:10.1029/96GL01360.

Johler, J. R., and L. C. Walters (1960), On the theory of reflection of low- and very-low-radiofrequency waves from the ionosphere, *Journal of Research NBS - Radio Propagation*, *64D*(3), 269–285.

Johnk, C. T. A. (1988), *Engineering electromagnetic fields and waves*, 2 ed., 637 pp., Wiley, Danvers, MA.

Kabirzadeh, R., R. A. Marshall, and U. S. Inan (2017), Early/fast VLF events produced by the quiescent heating of the lower ionosphere by thunderstorms, *Journal of Geophysical Research: Atmospheres*, *122*(12), 6217–6230, doi:10.1002/2017JD026528.

Kaiser, A. B. (1968), Identification of a New Type of Mode Conversion Interference in VLF Propagation, *Radio Science*, *3*(6), 545–549, doi:10.1002/rds196836545.

Kane, J. A., and J. Troim (1967), Rocket measurements of D-region electron number densities at sunrise, *Journal of Geophysical Research*, *72*(3), 1118–1120, doi:10.1029/JZ072i003p01118.

Kingma, D., and L. Ba (2015), Adam: A Method for Stochastic Optimization, in *ICLR*.

Koh, K. L., Z. Liu, and M. Füllekrug (2018), Lower ionosphere effects on narrowband VLF transmission propagation: fast variabilities and frequency dependence, *Radio Science*, doi:10.1002/2017RS006456.

Krider, E. P., R. C. Noggle, M. A. Uman, E. P. Krider, R. C. Noggle, and M. A. Uman (1976), A Gated, Wideband Magnetic Direction Finder for Lightning Return Strokes, *Journal of Applied Meteorology*, *15*(3), 301–306, doi:https://doi.org/10.1175/1520-0450(1976)015⟨0301:AGWMDF⟩2.0.CO;2.

Le Vine, D. M. (1987), Review of measurements of the RF spectrum of radiation from lightning, *Meteorology and Atmospheric Physics*, *37*(3), 195–204, doi:10.1007/BF01042441.

Lehtinen, N. G., and U. S. Inan (2008), Radiation of ELF/VLF waves by harmonically varying currents into a stratified ionosphere with application to radiation by a modulated electrojet, *Journal of Geophysical Research*, *113*, doi:10.1029/2007JA012911.

Liu, C., and S. Heckman (2011), The application of total lightning detection and cell tracking for severe weather prediction, in *91st American Meteorological Society Annual Meeting*, pp. 1–10.

Mach, D. M., D. R. MacGorman, W. David Rust, R. T. Arnold, D. M. Mach, D. R. MacGorman, W. D. Rust, and R. T. Arnold (1986), Site Errors and Detection Efficiency in a Magnetic Direction-Finder Network for Locating Lightning Strikes to Ground, *Journal of Atmospheric and Oceanic Technology*, *3*(1), 67–74, doi:10.1175/1520-0426(1986)003⟨0067:SEADEI⟩2.0.CO;2.

Marshall, R. A., and U. S. Inan (2010), Two-dimensional frequency domain modeling of lightning EMP-induced perturbations to VLF transmitter signals, *Journal of Geophysical Research: Space Physics*, *115*(A6), doi:10.1029/2009JA014761.

Marshall, R. A., U. S. Inan, and W. A. Lyons (2006), On the association of early/fast very low frequency perturbations with sprites and rare examples of VLF backscatter, *Journal of Geophysical Research*, *111*(D19), D19,108, doi:10.1029/2006JD007219.

Marshall, R. A., U. S. Inan, and T. W. Chevalier (2008), Early VLF perturbations caused by lightning EMP-driven dissociative attachment, *Geophysical Research Letters*, *35*(21), doi:10.1029/2008GL035358.

Martin, A., A. Ashish, B. Paul, B. Eugene, C. Zhifeng, C. Craig, S. C. Greg, D. Andy, D. Jeffrey, D. Matthieu, G. Sanjay, G. Ian, H. Andrew, I. Geoffrey, I. Michael, Y. Jia, J. Rafal, K. Lukasz, K. Manjunath, L. Josh, M. Dan, M. Rajat, M. Sherry, M. Derek, O. Chris, S. Mike, S. Jonathon, S. Benoit, S. Ilya, T. Kunal, T. Paul, V. Vincent, V. Vijay, V. Fernanda, V. Oriol, W. Pete, W. Martin, W. Martin, Y. Yuan, and Z. Xiaoqiang (2015), TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/.

Mathews, J., J. Breakall, and S. Ganguly (1982), The measurement of diurnal variations of electron concentration in the 60100 km ionosphere at Arecibo, *Journal of Atmospheric and Terrestrial Physics*, *44*(5), 441–448, doi:10.1016/0021-9169(82)90050-2.

Maxworth, A. S., M. Gołkowski, M. B. Cohen, R. C. Moore, H. T. Chorsi, S. D. Gedney, and R. Jacobs (2015), Multistation observations of the azimuth, polarization, and frequency dependence of ELF/VLF waves generated by electrojet modulation, *Radio Science*, *50*(10), 1008–1026, doi:10.1002/2015RS005683.

McCormick, J. C., M. B. Cohen, N. C. Gross, and R. K. Said (2018), Spatial and temporal ionospheric monitoring using broadband sferic measurements, *Journal of Geophysical Research: Space Physics*, *123*(4), 3111–3130, doi:10.1002/2017JA024291.

McRae, W. M., and N. R. Thomson (2000), VLF phase and amplitude: daytime ionospheric parameters, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(7), 609–618, doi:10.1016/S1364-6826(00)00027-4.

McRae, W. M., and N. R. Thomson (2004), Solar flare induced ionospheric D-region enhancements from VLF phase and amplitude observations, *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(1), 77–87, doi:10.1016/J.JASTP.2003.09.009.

Mika, Á., and C. Haldoupis (2008), VLF Studies During TLE Occurrences in Europe: ASummary of New Findings, *Space Science Reviews*, *137*(1-4), 489–510, doi:10.1007/s11214-008-9382-8.

Mika, A., C. Haldoupis, T. Neubert, H. T. Su, R. R. Hsu, R. J. Steiner, and R. A. Marshall (2006), Early VLF perturbations observed in association with elves, *Annales Geophysicae*, *24*(8), 2179–2189, doi:10.5194/angeo-24-2179-2006.

Mitchell, M. (2015), Very low frequency remote sensing of the lower ionosphere, Ph.D. thesis, University of Florida.

Mitra, A. P. (1974), *Ionospheric Effects of Solar Flares*, *Astrophysics and Space Science Library*, vol. 46, 307 pp., Springer Netherlands, Dordrecht, doi:10.1007/978-94-010-2231-6.

Miyake, Y., M. Ishii, T. Kawamura, and N. Honma (1995), On The Site Error of Wide Band Magnetic Direction Finder, *IEEJ Transactions on Power and Energy*, *115*(7), 825–831, doi:10.1541/ieejpes1990.115.7_825.

Moore, R. C., C. P. BarringtonLeigh, U. S. Inan, and T. F. Bell (2003), Early/fast VLF events produced by electron density changes associated with sprite halos, *Journal of Geophysical Research*, *108*(A10), doi:10.1029/2002JA009816.

Morelli, M., and U. Mengali (1999), Joint frequency and timing recovery for MSK-type modulation, *IEEE Transactions on Communications*, *47*(6), 938–946, doi:10.1109/26.771350.

Morfitt, D. G. (1977), Effective electron density distributions describing VLF/ELF propagation data, *Tech. rep.*, Naval Ocean Systems Center, Springfield, VA.

Morfitt, D. G., and C. H. Shellman (1976), 'MODESRCH', an improved computer program for obtaining ELF/VLF/LF mode constants in an earth-ionosphere waveguide, *Tech. rep.*, Defense Nuclear Agency, San Diego, CA.

Morgan, R. R. (1968), World-wide VLF effective-conductivity map, *Tech. rep.*, National Technical Information Service, Department of Commerce, Washington, D.C.

NaitAmor, S., M. A. AlAbdoadaim, M. B. Cohen, B. R. T. Cotts, S. Soula, O. Chanrion, T. Neubert, and T. Abdelatif (2010), VLF observations of ionospheric disturbances in association with TLEs from the EuroSprite-2007 campaign, *Journal of Geophysical Research: Space Physics*, *115*(A7), doi:10.1029/2009JA015026.

NaitAmor, S., M. B. Cohen, B. R. T. Cotts, H. Ghalila, M. A. AlAbdoadaim, and K. Graf (2013), Characteristics of long recovery early VLF events observed by the North African AWESOME Network, *Journal of Geophysical Research: Space Physics*, *118*(8), 5215–5222, doi:10.1002/jgra.50448.

Němec, F., M. Parrot, and O. Santolík (2015), Power line harmonic radiation observed by the DEMETER spacecraft at 50/60Hz and low harmonics, *Journal of Geophysical Research: Space Physics*, *120*(10), 8954–8967, doi:10.1002/2015JA021682.

Nicolet, M., and A. C. Aikin (1960), The formation of the D region of the ionosphere, *Journal of Geophysical Research*, *65*(5), 1469–1483, doi:10.1029/JZ065i005p01469.

Nina, A., V. Čadež, V. Srećković, and D. Šulić (2012), Altitude distribution of electron concentration in ionospheric D-region in presence of time-varying solar radiation flux, *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *279*, 110–113, doi:10.1016/J.NIMB.2011.10.019.

Pappert, R. A., and J. A. Ferguson (1986), VLF/LF mode conversion model calculations for air to air transmissions in the earth-ionosphere waveguide, *Radio Science*, *21*(4), 551–558, doi:10.1029/RS021i004p00551.

Pappert, R. A., and L. R. Hitney (1982), A WKB Program for ELF/VLF Earth-Ionosphere Excitation by Sources at Satellite Heights, *Tech. rep.*, Naval Ocean Systems Center, San Diego, CA.

Pappert, R. A., and L. R. Hitney (1988), Empirical modeling of nighttime easterly and westerly VLF propagation in the Earth-ionosphere waveguide, *Radio Science*, *23*(4), 599–611, doi:10.1029/RS023i004p00599.

Pappert, R. A., and D. G. Morfitt (1975), Theoretical and experimental sunrise mode conversion results at VLF, *Radio Science*, *10*(5), 537–546, doi:10.1029/RS010i005p00537.

Pappert, R. A., and L. R. Shockey (1971), WKB Mode Summing Program for VLF/ELF Antennas of Arbitrary Length, Shape and Elevation., *Tech. rep.*, Naval Electronics Lab Center, San Diego, CA.

Pappert, R. A., and F. P. Snyder (1972), Some Results of a Mode-Conversion Program for VLF, *Radio Science*, *7*(10), 913–923, doi:10.1029/RS007i010p00913.

Pappert, R. A., E. E. Gossard, and I. J. Rothmuller (1967), A Numerical Investigation of Classical Approximations Used in VLF Propagation, *Radio Science*, *2*(4), 387–400, doi:10.1002/rds196724387.

Pasko, V. P., U. S. Inan, and T. F. Bell (1996), Blue jets produced by quasi-electrostatic pre-discharge thundercloud fields, *Geophysical Research Letters*, *23*(3), 301–304, doi:10.1029/96GL00149.

Passi, R. M., and R. E. López (1989), A parametric estimation of systematic errors in networks of magnetic direction finders, *Journal of Geophysical Research*, *94*(D11), 13,319, doi:10.1029/JD094iD11p13319.

Pasupathy, S. (1979), Minimum shift keying: A spectrally efficient modulation, *IEEE Communications Magazine*, *17*(4), 14–22, doi:10.1109/MCOM.1979.1089999.

Peter, W. B., and U. S. Inan (2004), On the occurrence and spatial extent of electron precipitation induced by oblique nonducted whistler waves, *Journal of Geophysical Research*, *109*(A12), doi:10.1029/2004JA010412.

Raghuram, R., R. Smith, and T. Bell (1974), VLF Antarctic antenna: Impedance and efficiency, *IEEE Transactions on Antennas and Propagation*, *22*(2), 334–338, doi:10.1109/TAP.1974.1140777.

Rakov, V. A. (2016), *Fundamentals of Lightning*, Cambridge University Press, Cambridge, doi:10.1017/CBO9781139680370.

Ratcliffe, J. A. (1959), *The magneto-ionic theory and its applications to the ionosphere*, 206 pp., Cambridge University Press, Cambridge.

Raulin, J.-P., G. Trottet, M. Kretzschmar, E. L. Macotela, A. Pacini, F. C. P. Bertoni, and I. E. Dammasch (2013), Response of the low ionosphere to X-ray and Lyman-α solar flare emissions, *Journal of Geophysical Research: Space Physics*, *118*(1), 570–575, doi:10.1029/2012JA017916.

Richter, J. H. (1966), Application of Conformal Mapping to Earth-Flattening Procedures in Radio Propagation Problems, *Radio Science*, *1*(12), 1435–1438, doi:10.1002/rds19661121435.

Rodger, C. J., J. B. Brundell, R. H. Holzworth, and E. H. Lay (2009), Growing Detection Efficiency of the World Wide Lightning Location Network, in *AIP Conference Proceedings*, vol. 1118, pp. 15–20, AIP, doi:10.1063/1.3137706.

Rowe, J., A. Mitra, A. Ferraro, and H. Lee (1974), An experimental and theoretical study of the D-regionII. A semi-empirical model for mid-latitude D-region, *Journal of Atmospheric and Terrestrial Physics*, *36*(5), 755–785, doi:10.1016/0021-9169(74)90023-3.

Said, R. K. (2009), Accurate and efficient long-range lightning geo-location using a VLF radio atmospheric waveform bank, Ph.D. thesis, Stanford University.

Said, R. K., U. S. Inan, and K. L. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, *Journal of Geophysical Research*, *115*(D23), D23,108, doi:10.1029/2010JD013863.

Said, R. K., M. B. Cohen, and U. S. Inan (2013), Highly intense lightning over the oceans: Estimated peak currents from global GLD360 observations, *Journal of Geophysical Research: Atmospheres*, *118*(13), 6905–6915, doi:10.1002/jgrd.50508.

Salut, M. M., M. B. Cohen, M. A. M. Ali, K. L. Graf, B. R. T. Cotts, and S. Kumar (2013), On the relationship between lightning peak current and Early VLF perturbations, *Journal of Geophysical Research: Space Physics*, *118*(11), 7272–7282, doi:10.1002/2013JA019087.

Scherliess, L., R. W. Schunk, L. C. Gardner, J. V. Eccles, L. Zhu, and J. J. Sojka (2017), The USU-GAIM-FP data assimilation model for ionospheric specifications and forecasts, in *2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, p. 4, IEEE, doi:10.23919/URSIGASS.2017.8104978.

Sechrist, C. F. (1974), Comparisons of techniques for measurement of D-region electron densities, *Radio Science*, *9*(2), 137–149, doi:10.1029/RS009i002p00137.

Seddon, J. C. (1958), Differential absorption in the D and lower E regions, *Journal of Geophysical Research*, *63*(1), 209–216, doi:10.1029/JZ063i001p00209.

Shafer, D. C. (1994), Spread-spectrum VLF remote sensing of the ionosphere, Ph.D. thesis, Stanford.

Sheddy, C. H. (1968), A General Analytic Solution for Reflection From a Sharply Bounded Anisotropic Ionosphere, *Radio Science*, *3*(8), 792–795, doi:10.1002/rds196838792.

Shellman, C. H. (1986), A New Version of MODESRCH using Interpolated Values of the Magnetoionic Reflection Coefficients, *Tech. rep.*, Naval Ocean Systems Center, San Diego, CA.

Silber, I., C. Price, E. Galanti, and A. Shuval (2015), Anomalously strong vertical magnetic fields from distant ELF/VLF sources, *Journal of Geophysical Research: Space Physics*, *120*(7), 6036–6044, doi:10.1002/2015JA021141.

Singh, A. K., A. K. Singh, R. Singh, and R. P. Singh (2014), Solar flare induced D-region ionospheric perturbations evaluated from VLF measurements, *Astrophysics and Space Science*, *350*(1), 1–9, doi:10.1007/s10509-013-1699-4.

Smith, A. N., and J. C. Hanselman (1978), OMEGA Norway antenna system characteristics: modification and validation tests, *Tech. rep.*, Naval Ocean Systems Center, San Diego.

Smith, D. A., M. J. Heavner, A. R. Jacobson, X. M. Shao, R. S. Massey, R. J. Sheldon, and K. C. Wiens (2004), A method for determining intracloud lightning and ionospheric heights from VLF/LF electric field records, *Radio Science*, *39*(1), doi:10.1029/2002RS002790.

Swanson, E. (1983), Omega, *Proceedings of the IEEE*, *71*(10), 1140–1155, doi:10.1109/PROC.1983.12743.

Swanson, E. R. (1974), ELF-VLF applications in navigation and communications, in *ELF-VLF radio wave propagation; Proceedings of the Advanced Study Institute*, vol. 10, pp. 371–384, Spatind, Norway.

Taylor, W. L. (1960), VLF attenuation for east-west and west-east daytime propagation using atmospherics, *Journal of Geophysical Research*, *65*(7), 1933–1938, doi:10.1029/JZ065i007p01933.

Thébault, E., C. C. Finlay, C. D. Beggan, P. Alken, J. Aubert, O. Barrois, F. Bertrand, T. Bondar, A. Boness, L. Brocco, E. Canet, A. Chambodut, A. Chulliat, P. Coïsson, F. Civet, A. Du, A. Fournier, I. Fratter, N. Gillet, B. Hamilton, M. Hamoudi, G. Hulot, T. Jager, M. Korte, W. Kuang, X. Lalanne, B. Langlais, J.-M. Léger, V. Lesur, F. J. Lowes, S. Macmillan, M. Mandea, C. Manoj, S. Maus, N. Olsen, V. Petrov, V. Ridley, M. Rother, T. J. Sabaka, D. Saturnino, R. Schachtschneider, O. Sirol, A. Tangborn, A. Thomson, L. Tøffner-Clausen, P. Vigneron, I. Wardinski, and T. Zvereva (2015), International Geomagnetic Reference Field: the 12th generation, *Earth, Planets and Space*, *67*(1), 79, doi:10.1186/s40623-015-0228-9.

Thevenot, M., J.-P. Berenger, T. Monediere, and F. Jecko (1999), A FDTD scheme for the computation of VLF-LF propagation in the anisotropic earth-ionosphere waveguide, *Annals of Telecommunications*, *54*(297-310), 297–310.

Thiel, D., and R. Mittra (1997), Surface impedance modeling using the finite-difference time-domain method, *IEEE Transactions on Geoscience and Remote Sensing*, *35*(5), 1350–1356, doi:10.1109/36.628800.

Thomson, N. (1993), Experimental daytime VLF ionospheric parameters, *Journal of Atmospheric and Terrestrial Physics*, *55*(2), 173–184, doi:10.1016/0021-9169(93)90122-F.

Thomson, N. R. (1975), Whistler-mode signals: Group delay by crosscorrelation, *Geophysical Research Letters*, *2*(10), 451–452, doi:10.1029/GL002i010p00451.

Thomson, N. R. (1981), Whistler mode signals: Spectrographic group delays, *Journal of Geophysical Research: Space Physics*, *86*(A6), 4795–4802, doi:10.1029/JA086iA06p04795.

Thomson, N. R. (2010), Daytime tropical D region parameters from short path VLF phase and amplitude, *Journal of Geophysical Research: Space Physics*, *115*(A9), doi:10.1029/2010JA015355.

Thomson, N. R., and M. A. Clilverd (2000), Solar cycle changes in daytime VLF subionospheric attenuation, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(7), 601–608, doi:10.1016/S1364-6826(00)00026-2.

Thomson, N. R., and M. A. Clilverd (2001), Solar flare induced ionospheric D-region enhancements from VLF amplitude observations, *Journal of Atmospheric and Solar-Terrestrial Physics*, *63*(16), 1729–1737, doi:10.1016/S1364-6826(01)00048-7.

Thomson, N. R., M. A. Clilverd, and W. M. McRae (2007), Nighttime ionospheric D region parameters from VLF phase and amplitude, *Journal of Geophysical Research: Space Physics*, *112*(A7), doi:10.1029/2007JA012271.

Thomson, N. R., C. J. Rodger, and M. A. Clilverd (2011), Daytime D region parameters from long-path VLF phase and amplitude, *Journal of Geophysical Research: Space Physics*, *116*(A11), doi:10.1029/2011JA016910.

Thomson, N. R., C. J. Rodger, and M. A. Clilverd (2012), Tropical daytime lower D-region dependence on sunspot number, *Journal of Geophysical Research: Space Physics*, *117*(A10), doi:10.1029/2012JA018077.

Thomson, N. R., M. A. Clilverd, and C. J. Rodger (2014), Low-latitude ionospheric D region dependence on solar zenith angle, *Journal of Geophysical Research: Space Physics*, *119*(8), 6865–6875, doi:10.1002/2014JA020299.

Thomson, N. R., M. A. Clilverd, and C. J. Rodger (2017), Midlatitude ionospheric D region: Height, sharpness, and solar zenith angle, *Journal of Geophysical Research: Space Physics*, *122*(8), 8933–8946, doi:10.1002/2017JA024455.

Tsurutani, B. T., O. P. Verkhoglyadova, A. J. Mannucci, G. S. Lakhina, G. Li, and G. P. Zank (2009), A brief review of solar flare effects on the ionosphere, *Radio Science*, *44*(1), doi:10.1029/2008RS004029.

Voss, H. D., W. L. Imhof, M. Walt, J. Mobilia, E. E. Gaines, J. B. Reagan, U. S. Inan, R. A. Helliwell, D. L. Carpenter, J. P. Katsufrakis, and H. C. Chang (1984), Lightning-induced electron precipitation, *Nature*, *312*(5996), 740–742, doi:10.1038/312740a0.

Wait, J. R. (1968a), Mode conversion and refraction effects in the Earth-ionosphere waveguide for VLF radio waves, *Journal of Geophysical Research*, *73*(11), 3537–3548, doi:10.1029/JA073i011p03537.

Wait, J. R. (1968b), On the theory of VLF propagation for a step model of the nonuniform earthionosphere wave guide, *Canadian Journal of Physics*, *46*(17), 1979–1983, doi:10.1139/p68-539.

Wait, J. R. (1970), *Electromagnetic waves in stratified media*, 608 pp., Pergamon Press.

Wait, J. R., and K. P. Spies (1964), Characteristics of the earth-ionosphere waveguide for VLF radio waves, *Tech. rep.*, National Bureau of Standards, Boulder, CO.

Wait, J. R., and K. P. Spies (1965), Influence of Finite Ground Conductivity on the Propagation of VLF Radio Waves, *Radio Science Journal of Research NBS/USNC-URSI*, *69D*(10).

Wait, J. R., and L. C. Walters (1963a), Reflection of VLF Radio Waves From an Inhomogeneous Ionosphere. Part I. Exponentially Varying Isotropic Model, *Journal of Research of the National Bureau of Standards*, *67*(3).

Wait, J. R., and L. C. Walters (1963b), Reflection of VLF Radio Waves From an Inhomogeneous Ionosphere. Part II. Perturbed Exponential Model, *Journal of Research of the National Bureau of Standards*, *670*(5).

Wait, J. R., and L. C. Walters (1963c), Reflection of VLF Radio Waves From an Inhomogeneous Ionosphere. Part III. Exponential Model With Hyperbolic Transition, *Journal of Research of the National Bureau of Standards*, *67*(6).

Walker, D. (1965), Phase Steps and Amplitude Fading of VLF Signals at Dawn and Dusk, *Radio Science Journal of Research*, *69*(11), 1435–1443.

Watt, A. D. (1967), *VLF radio engineering*, 703 pp., Pergamon Press, Oxford.

Wessel, P., and J. M. Becker (2008), Interpolation using a generalized Green's function for a spherical surface spline in tension, *Geophysical Journal International*, *174*(1), 21–28, doi:10.1111/j.1365-246X.2008.03829.x.

Whitten, R. C., and I. G. Poppoff (1971), *Fundamentals of Aeronomy*, 446 pp., J. Wiley.

Wood, T. G., and U. S. Inan (2004), Localization of individual lightning discharges via directional and temporal triangulation of sferic measurements at two distant sites, *Journal of Geophysical Research: Atmospheres*, *109*(D21), doi:10.1029/2004JD005204.

Yamashita, M., and K. Sao (1974a), Some considerations of the polarization error in direction finding of atmospherics-I. Effect of the Earth's magnetic field, *Journal of Atmospheric and Terrestrial Physics*, *36*(10), 1623–1632, doi:10.1016/0021-9169(74)90200-1.

Yamashita, M., and K. Sao (1974b), Some considerations of the polarization error in direction finding of atmospherics-II. Effect of the inclined electric dipole, *Journal of Atmospheric and Terrestrial Physics*, *36*(10), 1633–1641, doi:10.1016/0021-9169(74)90201-3.

Zimmermann, H.-G., A. Minin, and V. Kusherbaeva (2011), Comparison of the Complex Valued and Real Valued Neural Networks Trained with Gradient Descent and Random Search Algorithms, in *ESANN*.

Zoghzoghy, F. G. (2015), Statistical analysis and modeling lightning using radio remote sensing, Ph.D. thesis, Stanford University.